The Persistent Effect of Gender Division of Labour: African American Women After Slavery

Anna Baiardi*

November 16, 2016

Abstract

This paper explores the role of historic gender division of labour in shaping gender norms. To answer this question, I analyse whether differences in the gender division of labour during slavery have a persistent effect on African American women's labour market outcomes after the end of slavery. I use variation in the production of cotton and tobacco across counties during slavery as a proxy for gender division of labour: tobacco was characterized by a starker gender division of labour compared to cotton. Using data from 1870 to 2010, I show that women living in counties with lower degrees of gender division of labour (higher cotton production relative to tobacco) are more likely to participate in the labour market and have higher occupation income scores, for at least 70 years after emancipation. Furthermore, to disentangle gender roles from local labour demand effects, I analyse the labour force participation of migrants from counties with high historic cotton and tobacco production who relocated to urban areas. The results indicate that gender roles affect women's labour supply, which in turn affects their labour market experience, resulting in differences in occupation income score.

Keywords: Gender roles, labour market, slavery, historical persistence.

JEL Classification: D03, J16, J21, N31.

*University of Warwick. Email: a.baiardi@warwick.ac.uk. I thank Sascha Becker, Clément Imbert, Rocco Macchiavello, Luigi Pascali, Roland Rathelot, and Chris Woodruff for their valuable advice, and the participants of the following workshops: Warwick CAGE, Namur Poder Summer School 2016, Oxford Development 2016. I gratefully ackowledge the financial support of the ESRC.

1 Introduction

This study relates to the origins and consequences of differences in gender roles, i.e. the beliefs and norms dictating the type of behaviour which is generally accepted in the society for women and men. Recent work in economics has explored both the possible mechanisms behind the origins of gender roles and their effects. Differences in gender roles can affect women's labour market outcomes, by influencing their labour force participation decision (Alesina et al., 2013; Fernández, 2013). Moreover, if these cultural beliefs can be transmitted through generations, they can lead to path dependence (Farré and Vella, 2013). I add to this literature by analysing whether slavery, by establishing agricultural practices which promoted differences in gender divison of labour, played a role in the formation of gender norms among the African American population. I then explore the persistence of these mechanisms after slavery in regard to the labour market.

I test this hypothesis by exploiting differences in the prevalence of cotton and tobacco in different geographic areas, as the cultivation of these crops corresponded to different degrees of gender division of labour during slavery. Cotton cultivation in the U.S. South was characterised by female and male slaves working side by side. This is opposed to tobacco plantations: the decline in production of the crop during the 19th century reduced the need for female labour in agriculture, which resulted in a sharper gender division of labour, with men working in agriculture and women mostly performing domestic work. I exploit variation in cotton prevalence relative to tobacco during slavery in order to explain differences in labour market outcomes of African American women between 1870 - five years after emancipation - and 2010. I use U.S. census data at the individual level, which allows me to control for individual characteristics and county-specific effects.

My findings indicate that gender division of labour during slavery had a positive and persistent effect on labour force participation and the occupation income score of African American women. The results are consistent with differences in the gender division of labour due to the prevalence of cotton and tobacco affecting the decision of women to participate in the labour force after slavery. The increase in labour market experience of women in cotton regions in turn influenced their ability to find jobs paying higher wages. The effects decline in magnitude over time but persist for at least 70 years after emancipation.

I firstly analyse the effect of cotton and tobacco cultivation on the labour market outcomes of African American women. The main explanatory variable of interest is a

county-level measure of cotton production relative to tobacco during slavery, which I interact with a dummy variable indicating whether the individual is female. In areas where cotton was more prevalent than tobacco, African American women were more likely to participate in the labour market and to have higher occupation income scores; this implies that, conditional on labour force participation, women in cotton regions were more likely to be employed in occupations paying higher wages. The effects are economically significant: in 1880, a one standard deviation increase in production of cotton relative to tobacco increases labour force participation of women relative to men by 8 percentage points, and occupation income scores by about 15 percent. These results are robust to using different measures of prevalence of the two crops during slavery. In particular, because production may be subject to measurement error due to changes in production over time, and is potentially affected by pre-existing gender roles, I use an instrumental variable approach: I exploit the differences in production predicted by differences in the suitability of the land for the cultivation of the two crops. Moreover, to reduce the concern that differences in income may be driving the results, I control for proxies of family income.

Cotton and tobacco were prevalent in different geographical regions: cotton was more widespread in the Deep South, whereas tobacco was cultivated in the Upper South. Thus the presence of omitted variables, such as differences in local institutions and labour market conditions, may result in biased estimates. I address this concern by analysing the labour market outcomes of migrants originating from slave states who relocated to urban areas during the Great Migration: I compare migrants from counties with high production of cotton and tobacco during slavery to migrants originating from counties with higher prevalence of other crops. Because individual-level data for county of origin is not available from the early 20th century censuses, I construct a data set of migrants by using the fact that a significant proportion of African Americans after emancipation adopted the surname of the plantation owners they worked for. I infer the origin of African Americans who have migrated after slavery by matching their mother's surname with the surnames of plantation owners in a county with high prevalence of a particular crop. I focus on mothers' rather than individuals' surnames to limit the measurement error arising from women adopting their husband's surname after marriage. To address possible differences in selection into migration across regions, I compare women and men originating from areas with similar levels of crop prevalence; moreover, I control for individual characteristics and allow their effect to vary for women and men; in addition, I analyse whether migrants originating from different regions have different observable characteristics. I conduct this

analysis for year 1930 and I find that migrants from cotton counties are more likely to participate in the labour force compared to other migrants. These results are confirmed using data on migration from the 1940 census.

I then investigate the mechanisms driving the higher occupation income scores of women in cotton regions. Using data from 1880, I find evidence to suggest differences in labour market experience: young women in cotton-prevalent counties are more likely to participate in the labour force compared to their counterparts in tobacco-prevalent counties, and as their age increases, they are more likely to have higher occupation income scores conditional of being in the labour market. This suggests that women in the former regions are likely to accumulate more labour market experience. Moreover, I investigate the extent to which gender roles and labour market experience can be transmitted through generations, by testing whether the labour market status of mothers can explain differences in labour force participation of their daugthers and their occupation income scores. Daughters whose mother participates in the labour force are more likely to work themselves. Moreover, women whose mother participates in the labour force have higher occupation income score in areas with higher relative cotton prevalence, suggesting that the higher labour market experience of mothers in cotton regions positively affects their daughter's probability of finding jobs paying higher wages.

I next explore alternative mechanisms which may contribute to the results. Firstly, I investigate differences in labour demand for agriculture: a larger proportion of African American women are employed in agriculture in counties where cotton prevalence is higher; however, labour market outcomes of other ethnic groups do not appear to exhibit the same patterns. Moreover, conditional on working in service and manufacturing, women in cotton regions have better occupations (in terms of wages) compared to women in other areas. While these results do not rule out differences in labour demand, they suggest that other mechanisms may have an effect. I show evidence for two other channels: lower discrimination towards African American women in cotton regions and better access to social networks. However, none of these mechanisms alone can explain the entire increase in labour force participation or occupation income score.

This work relates to several strands of literature. Firstly, it contributes to the literature on the persistence of gender roles. Alesina et al. (2013) are the first to document a cross country relation between current gender roles and the historic use of the plough, which transformed agriculture into a male-dominated activity. I address this question in the context of slavery for two reasons. Firstly, it can shed light on the long lasting conse-

quences of slavery for African American women; secondly, in this setting the endogeneity issue due to selective settlements based on pre-existing gender roles is mitigated, given the limited choice that slaves had over their location. This work also relates to Farré and Vella (2013), who study the intergenerational transmission of gender roles from mothers to their children. Fernández (2013) instead explains changes in female labour force participation over time in the U.S. by proposing a model in which women learn about the long-term payoffs of working from older generations.

This paper also relates to the literature analysing the long term consequences of slavery on labour market outcomes of African Americans, in particular women. With historical census data, Goldin (1977) was the first to document that forced labour during slavery could be one of the causes underlying the differences in labour force participation of African American and white women in the U.S. Up until 1980s, the percentage of black women in the labour force was consistently higher than that of white women (Boustan and Collins, 2014), and the difference survives even after controlling for income Goldin (1977, 1991). This suggests that poverty is not the only cause of higher female labour force participation rates in the African American community, but that other factors such as preferences may play a role. Goldin (1977) outlines the hypothesis that the "stigma" associated with being employed outside the house, prevalent within white communities until the mid of the 20th century, may not affect African Americans. She relates these differences to slavery: forced labour caused women and men to work side by side, therefore hindering the formation of cultural norms related to gender roles. This could have resulted in black women being more likely to work than white women, at the same level of income. Boustan and Collins (2014) test this idea with U.S. historic census data and find that black women whose mother is in the labour force are more likely to participate in the labour force, and even more so if the mother was born in a Southern state. This paper contributes to the literature by exploring the hypothesis that different practices regarding gender division of labour during slavery may explain differences in labour market outcomes within the African American community rather than focusing on the racial gap, analysing both labour market status and the type of work they perform.

Other studies focus on the long term consequences of slavery in the U.S. Ager et al. (2016) link slavery to labour force participation of African American women by studying the negative effect of an income shock - the Boll Weevil Plague, which affected cotton plantations in the early 20th century - on female labour supply. They find that African American women were particularly affected by this pest, and their results indicate that

cotton was a very important source of income for black women. Chay and Munshi (2013)'s study is relevant to this work as it suggests a positive relation between historical labour intensive plantations and the formation of networks in the post-slavery U.S. Their evidence demonstrates that networks played a role on migration patterns and voting behaviour of black communities after Emancipation. Moreover, Christian (2014) explores the long run effects of lynching on labour market outcomes of African Americans. In addition, this paper relates to the work of Nunn (2008) and Nunn and Wantchekon (2011) on the persistent effect of slavery on economic development and trust in the origin countries of slaves, of Bertocchi (2015) on the consequences of slavery both in African countries and in North and Latin America, and of Fenske et al. (2014) on the effect of the suppression of the slave trade on conflict in Africa.

The paper is organized as follows. In section 2 I summarise the historical background, describing the role of women in cotton and tobacco plantations during slavery. In section 3 I give a detailed description of the data used for the analysis, and in section 4 I outline the empirical strategy. Section 5 focuses on the main results, and in section 6 I investigate the mechanisms. Finally, section 7 concludes.

2 Historical background

2.1 Cotton and tobacco during slavery and the role of women

Cotton and tobacco have been cultivated in the U.S. since the beginning of the 17th century and were among the major crops cultivated in the U.S. during the slavery period.

Tobacco cultivation started in Virginia, in the Chesapeake Bay and spread to North Carolina, Kentucky, Pennsylvania and Tennessee, whereas cotton was mainly cultivated in the Deep South. The two crops were originally cultivated on small farms, but later large plantations emerged, and slavery played a major role in this transformation (Walsh, 1989). Tobacco was the main crop produced in the U.S. until the 18th century, when prices dropped due to overproduction, and export demand for the crop was greatly reduced. This caused either a shift towards more cotton production (mostly in Southern regions, more suitable for this crop), or diversification in the Chesapeake; the latter involved an increased production of subsistence crops in addition to tobacco (Walsh, 1989). This shift forced tobacco planters to change their production processes: as tobacco's importance as a cash crop decreased, there was less demand for female labour in the fields. This

resulted in males mostly working in the fields and women performing domestic work and other types of tasks. Moreover, as described by Walsh (1989), the change in production was brought about by the introduction of the plough, which may have contributed to the division of labour by gender.

The role of women in cotton cultivation during slavery has been described by several authors in the literature, in the fields of both economics and history. Metzer (1975), Shlomowitz (1979) and Jones (2009) suggest that women have a comparative advantage in picking cotton, the most labour intensive activity in cotton plantations, and Wayne (2007) describes the work in cotton fields, stating that women, men and children worked side by side. Goldin and Sokoloff (1984) suggest that women have a comparative advantage in crops that required extensive cultivation, such as cotton and tobacco. Whereas in principle cotton and tobacco are both cultivated on large plantations, as described above, some differences arose in the way the plantations were organized, which lead to tobacco cultivation promoting starker differences in the types of tasks performed by men and women.

Ethnographic studies have attempted to identify the origins of African Americans, based on data about slave voyages landing in different ports in the U.S. Walsh (2001) studies the origins of African Americans in the Chesapeake, using data from the *Virginia Slave Trade Statistics* and from the *Trans-Atlantic Slave Trade Database*. She suggests that most of the African Americans in the Upper Chesapeake in the 18th century came from the northern parts of the West African coast, approximately corresponding with the territory of Senegal and Ghana. In the lower Chesapeake, African Americans originated from more southern parts of Africa: eastern Nigeria and West Central Africa (Congo and Angola). It is estimated that the origin of slaves in South Carolina was similar to that of the lower Chesapeake, i.e. West Central Africa.

Slavery ended with the Emancipation Proclamation in 1863. While tobacco cultivation had already decreased dramatically during the 19th century, cotton cultivation declined after Emancipation. The drop in cotton production, which was caused by declining prices alongside the spreading of the Boll Weevil, is regarded by some as one of the main causes of the Great Migration (Higgs, 1976; Lange et al., 2009). This term refers to the mass migration of African Americans from the agricultural South to the North, in particular to Northern cities, between 1910 and 1970. Migration of African Americans from the South to the North was very limited before the 1910s, partly because of lack of demand for labour for African Americans, but also, as described by Ransom and Sutch (2001), as

a consequence of the low upward occupational mobility of Southern African Americans working in agriculture (Collins and Wanamaker, 2014).

2.2 Surname adoption after slavery

As slavery was eliminated, African Americans, who typically did not possess formal surnames during slavery, suddenly needed to adopt one. Those seeking employment, as well as conscripts in the army, were required to have a permanent full name. Although precise information about how African Americans chose their surnames is difficult to obtain, anecdotal evidence suggests that a significant percentage of African Americans decided to keep the surname of the slaveholder for whom they used to work. For instance Gutman et al. (1976), in his book "The Black Family in Slavery and Freedom, 1750-1925", which collects information about several families and descendants of ex-slaves, finds that 27% of the 181 individuals interviewed in South Carolina, and 36% of the 217 interviewed in Texas, claimed that their family retained the surnames of the last slaveholder. However, often those individuals who did not retain the surname of their last slaveholder decided to adopt the surname of a previous slaveholder; others would choose the last name of a benefactor or of another white person that they knew and liked. In fact, surnames of African Americans could not be distinguished from those of white people in the South.

Other popular choices of surnames among African Americans are those of ex-presidents: Johnson, Brown, Jefferson and Davis were among the most popular last names. Much less common was the habit of using the name of cities of towns, or the use of nicknames, occupational names (with the exception of Smith) and African names.

3 Data

In this paper I use cross-sectional data from historical and recent U.S. Censuses. I merge these with measures of crop production during slavery.

I use microdata from the U.S. population censuses from 1870 to 2000, and from the American Community Survey for 2010. The samples used range from 1% to 100% of the total samples of each year, depending on availability. I use microdata from the 10% sample of the 1880 Population Census to perform the analysis on 1880. The advantage of this data set is that all minorities are oversampled: it contains 1 in 5 observations for African Americans (and other minorities) of those contained in the 100% Census. It includes

individual-level information about age, sex, labour force status, detailed occupation and industry, occupation income score, marital status and family members. Occupation income scores are constructed by assigning a median earning to each occupation, measured in 1950 U.S. dollars. They have been used as a proxy for wages, since data about earnings is not available in censuses prior to 1940 (Biavaschi et al., 2013). I use the log of occupation income scores as one of the main dependent variables in my analysis.

Slaveholder surnames used to construct the data set of migrants are obtained from the 1860 sample of the slave schedules available from the Integrated Public Use Microdata Series (IPUMS). This is a 1-in-20 sample of the slaves enumerated in 1860, which includes the surname of slaveholders (the owners of the plantations where the slaves were registered), and limited information about the slaves themselves, including the total number of slaves in each holding.[1] All of the census data used in this paper are available from the IPUMS, North Atlantic Population Project (NAPP) or National Historical Geographic Information System (NHGIS) websites.

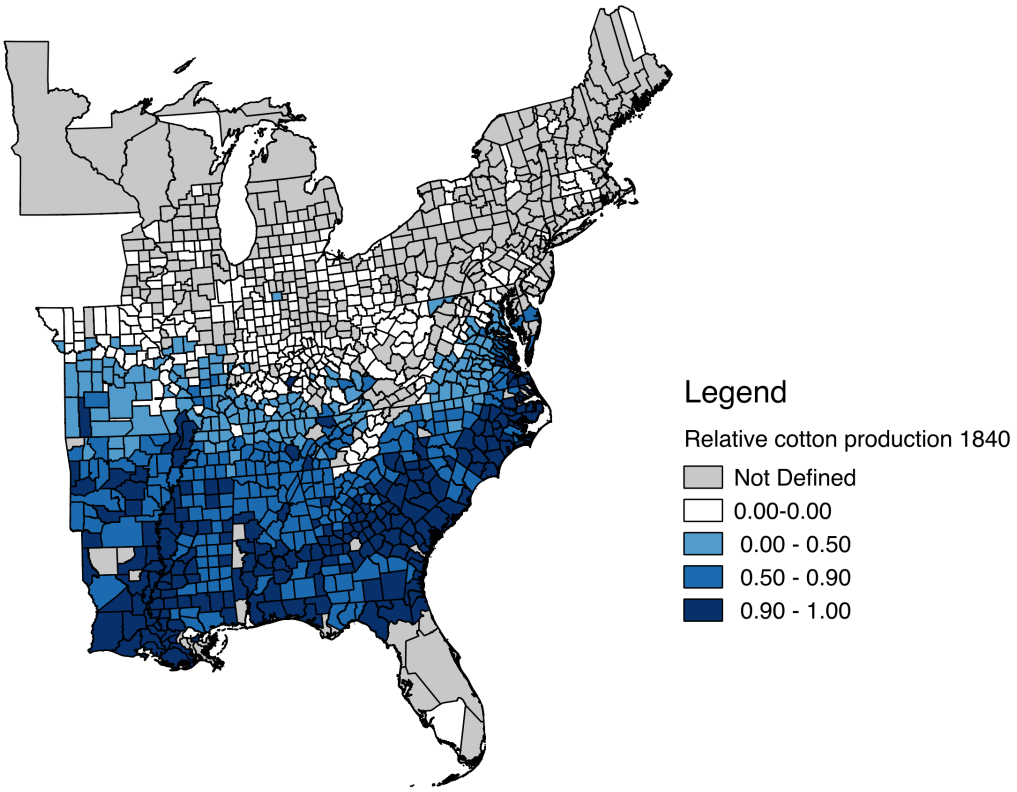
I obtain crop production and acreage variables at the county level from the Agriculture Census of years 1840 and 1880 respectively. To construct the instrument for crop production, I use detailed crop suitability data from Global Agro-Ecological Zones (GAEZ), which I aggregate at the county level to match the individual data from the Population Censuses. Crop suitability indices determine how well soil conditions match crop requirements given specific levels of inputs and irrigation. The suitability indices used in this paper are calculated for low input levels and rain-fed production (no irrigation). This is to reduce to a minimum the possibility that the variation in suitability actually reflects differences in management or irrigation systems. Cotton and tobacco suitability indices range from 0 to 8, where higher numbers indicate higher suitability.

3.1 Measures of gender division of labour

As I am interested in capturing variation in gender roles during slavery, the ideal measure would be the number of female slaves working in cotton and tobacco plantations across geographical regions in the U.S. To the best of my knowledge, this measure is not available. Therefore, I construct a proxy using production data for cotton and tobacco prevalence during slavery. The earliest county-level data is available for year 1840: it reports the total amount produced of different crops, measured in pounds.

[1] For a more detailed description of the slave schedules, see Ruggles et al. (2010).

Figure 1: Relative Cotton Production 1840



Notes: The map shows the main measure of cotton production relative to tobacco in 1840. Source: U.S. Agricultural Census 1840.

I construct an index of production of cotton relative to tobacco, which allows me to capture the intensity of the treatment, i.e. the level of gender division of labour during slavery in a specific county. I calculate this measure as follows:

$$relative\,cotton\,production\,1840_c = \frac{cotton\,1840_c}{cotton\,1840_c\,+\,tobacco\,1840_c}$$

where $cotton\,1840$ and $tobacco\,1840$ measure the pounds of cotton and tobacco produced in county c as recorded in the 1840 Agriculture Census.

The higher the index of relative production, the more cotton was produced relative to tobacco, which implies less gender division of labour overall in that county. Note that this measure is only defined for counties where total cotton and tobacco production is greater than zero, and therefore allows me to compare areas where at least one of the two crops

was grown. Figure 1 shows the geographic distribution of cotton production relative to tobacco: cotton cultivation was more prevalent in the South, whereas tobacco was mainly concentrated in the East.

The measure outlined above has some limitations. Particularly, it may be affected by productivity: one county may produce more cotton compared to tobacco not because the first employs more workers, but because labour is more productive. Although this is a theoretical possibility, by comparing figures 1 and C.1, which show relative crop production in 1840 and relative suitability respectively, production of the two crops crops appears to be geographically concentrated in regions where suitability for that specific crop is higher. I carry out additional robustness checks with a similar measure constructed with the number of acres of land cultivated with each crop in 1880, the first year for which this data is available at the county level. I construct this measure as follows:

$$relative\,cotton\,share\,1880_c = \frac{cotton\,acres\,1880_c}{cotton\,acres\,1880_c\,+\,tobacco\,acres\,1880_c}.$$

.

Figure C.2 shows cotton acreage relative to tobacco by county in 1880.

One further concern with the measures outlined above is that they ignore the extent of the production of cotton and tobacco relative to the total agricultural production. This would not be an issue if one assumes that all other crops are gender neutral. To avoid making this assumption, I run robustness checks using a similar measure for relative production, which computes the relative production of the two crops as a share of total crop production by county:

$$relative\,cotton\,prod.\,to\,total\,1840_c = \frac{cotton\,1840_c - tobacco\,1840_c}{total\,crop\,output\,1840_c}.$$

To ease interpretation as a share of total production, I apply a linear transformation by adding one and dividing by two, in order to obtain a variable taking values between 0 and 1.

Table 1 shows summary statistics for the different measures of relative cotton prevalence[2]. Notice that the measures differ in the number of counties they are identified for. The main 1840 measure is available for 804 counties.

[2]These include the instrument for cotton production relative to tobacco constructed using cotton and tobacco suitability, which I describe in section 4.2

Table 1: Relative Cotton Prevalence by County

	Mean	Median	SD	Min	Max	N Counties
Relative Cotton Production 1840	.491	.457	.466	0	1	804
Relative Cotton Share of Farmland 1880	.420	0	.484	0	1	1,736
Relative Cotton Production as a Share of Total Production 1840	.546	.5	.128	.133	1	1,094
Relative Cotton Suitability	.888	.982	.499	.125	2.738	2,999

Notes: Data from 1840 and 1880 Agricultural Censuses, and Gaez Crop Suitability. *SD* indicates standard deviation. The number of observations represents the number of U.S. counties for which information about crops is available.

3.2 Descriptive statistics 1880 sample

I use data from the 1880 population census to analyse the short term effects of the gender divison of labour.

Table 2: Descriptive Statistics - African Americans in 1880

	Relative Cotton Production>median			Relative Cotton Production <median			
	Mean	Max	Min	Mean	Max	Min	N
Female	.515	0	1	.497	0	1	346,288
Panel A: Women							
Labour Force Participation	.486	0	1	.298	0	1	161,376
Occupation Income Score (100$)	10.975	4	80	9.099	4	80	71,593
Age	35.763	25	54	36.089	25	54	176,887
Married	.725	0	1	.683	0	1	176,887
Urban	.142	0	1	.256	0	1	176,887
Literate	.167	0	1	.294	0	1	176,887
N Children	2.604	0	9	2.302	0	9	176,887
Panel B: Men							
Labour Force Participation	.995	0	1	.994	0	1	164,943
Occupation Income Score (100$)	15.225	3	80	16.380	3	80	164,174
Age	36.051	25	54	36.367	25	54	169,401
Married	.817	0	1	.739	0	1	169,401
Urban	.127	0	1	.253	0	1	169,401
Literate	.238	0	1	.363	0	1	169,401
N Children	2.358	0	9	1.997	0	9	169,401

Notes: I include individuals aged between 25 and 54, and split the sample into individuals who live in counties with *relative cotton production* below and above median. The statistics on occupation income score are conditional on labour force participation. Source: U.S. 1880 full Population Census.

Table 2 shows descriptive statistics of the sample of African Americans in 1880 of age between 25 and 54, in counties where cotton prevalence is higher than median and below median. I split the sample into males and females. A higher percentage of the African American population is female in counties where cotton prevalence is high. African American women in counties with higher cotton production during slavery have higher

labour force participation rates, higher occupation income scores, higher probability of being married, lower probability of living in urban areas and lower literacy rates. For men, labour force participation is almost universal. In contrast to the sample of women, African American men in counties with higher cotton production have slightly lower average occupation income scores.

Figure C.1 shows the top ten occupations by gender among the African American population. The vast majority - about 80% - of males are employed in agriculture, which is also the first industry by employment of women. However, a much larger percentage of females are employed in private households, as domestic help, compared to males: 35 versus 4 percent. The remaining share of employed individuals - about 15% of men and 6% of women - is employed in manufacturing, construction and other services.

Figure C.4 shows the evolution of the share of African Americans by gender employed in agriculture over time, with data from the decennial U.S. censuses. Notice that whereas agriculture was one of the main activities at the beginning of the 20th century, its importance reduces dramatically starting from 1930, and the decrease is relatively larger for women than it is for men.

4 Empirical strategy

4.1 OLS regression

To analyse the effect of cotton suitability on labour market outcomes of African American women I estimate the following equation, restricting the sample to African Americans:

$$\begin{aligned} y_{i,c} = \beta_0 + \beta_1 relative\,cotton\,prod_c \times female_i + \beta_2 female_i + \\ \beta_3 \mathbf{X}_i + \beta_4 \mathbf{X}_i \times female_i + \beta_5 \mathbf{Z}_c \times female_i + \theta_c + \epsilon_{i,c} \end{aligned} \tag{1}$$

where i is an African American individual; c indicates an American county; y is either a binary variable indicating whether the individual is in the labour force or a continuous variable measuring the log of occupation income score; *relative cotton prod* is the measure of relative prevalence of cotton and tobacco in 1840 as described in section 3.1; *female* is a binary variable taking value of 1 if the individual is a woman and 0 otherwise; $\boldsymbol{X}$ is a vector of individual controls (age, urban vs rural, marital status, income of spouse); $\boldsymbol{Z}$ is a vector of county controls fixed at year 1870 (number of manufacturing establishments, population, improved land in farms), which I interact with the *female*

dummy, and θ represents U.S. county fixed effects. I include state times $female$ fixed effects as a robustness check. I estimate this equation with cross sectional data from years 1870 to 2010, and all regressions are restricted to individuals between 25 and 54 years old, to facilitate comparisons with the existing literature (Goldin, 1977, 1990).

I choose the above as my main specification, as opposed to only analysing the effect for African American women across counties, because the interaction term allows to control for county specific effects. An alternative specification would compare African American women with white women. However, for the beginning of the time period considered, white women do not provide a good comparison group for black women, as black females were mostly employed in agriculture and private households, which were uncommon occupations among white women (Boustan and Collins, 2014). Nevertheless, for later years, particularly since the mid 20th century, it is interesting to compare black and white women's labour market outcomes. I analyse this in section 5.2.

4.2 Instrumental variable regression

There are two main concerns with the above empirical strategy. One may be concerned that cotton and tobacco production would be endogenous to cultural factors related to gender roles. Even though slaves could not take decisions about where to locate or what crops to grow, plantation owners may have had different preferences in relation to gender roles which in turn may have affected their decisions. Moreover, the OLS estimates could be biased due to measurement error: the prevalence of the two crops may have changed over time during slavery, and this cannot be accounted for using only the 1840 data. In addition, as described above, the measure may be affected by prices or productivity. Notice that the first concern would introduce a positive bias in the OLS estimates, whereas in the second case the bias would be towards zero.

Therefore, I run instrumental variable regressions where I instrument relative prevalence of cotton compared to tobacco with the relative suitability for the two crops.

I measure relative crop suitability by simply taking the ratio of the suitability for the two crops:

$$relative\, suitability_c = \frac{cotton\, suitability_c}{tobacco\, suitability_c}.$$

Figure C.1 shows the measure of relative cotton suitability by county. Notice that in the counties where African Americans were living in 1880 (see figure C.7), crop production

and suitability follow similar patterns.

The first stage regression is:

$$\begin{aligned} relative\,cotton\,prod_c \times female_i = \alpha_0 + \alpha_1 relative\,suitability_c \times female_i + \\ \alpha_2 female_i + \alpha_3 \mathbf{X}_i + \alpha_4 \mathbf{X}_i \times female_i + \alpha_5 \mathbf{Z}_c \times female_i + \delta_c + \epsilon_{i,c} \end{aligned} \tag{2}$$

where *cotton share* and *relative suitability* are the measures of relative cotton prevalence described in section 3.1 and δ represents county fixed effects.

While mitigating the issues mentioned above, instrumenting for crop production with crop suitability cannot address another possible source of endogeneity: the potential sorting of slaves across cotton and tobacco regions according to their pre-existing characteristics. For instance, it is possible that planters might have preferred slaves from cultures with different gender roles for the cultivation of different crops. I address this issue in section 2, where I describe historical and ethnographic evidence which finds no stark differences in the origin of slaves across states.

4.3 Identifying effect for migrants

To understand whether the effects of the gender division of labour on the labour market for African American women can be partly attributed to the shaping of gender roles in the form of attitudes towards the labour market, I compare labour market outcomes of African Americans who were born in areas where cotton was cultivated, but subsequently migrated to other areas. One major challenge is that information about migration is not available at the county level from the U.S. Census, except for the year 1940. However, since the main migration wave started in the 1910s, it is important to analyse migration for earlier years as well. To overcome the data limitations, I adopt two complementary identification strategies. The first approach implements surname matching of African Americans and white slaveholders and compares the geographic location of the former relative to the latter in order to obtain a measure of migration for 1930; the second exploits the migration information available in the 1940 census.

For the first strategy, I retrieve information about migration by making use of the fact that after emancipation it was common for African Americans to adopt the surnames of their former slaveholders (Van Deburg, 1997), as one was needed for official documents. Sharing the surname of a slaveholder that owned a plantation in a cotton (or tobacco)

county increases the likelihood of the individual having been a slave herself, or being a descendant of a slave who worked in a cotton (or tobacco) plantation. I obtain a proxy for migration by analysing the location of the individuals whose surname matches that of a former slaveholder: if the individual is located in a state where slavery was not prevalent, I treat her as a migrant.[3] However, an additional complication arises when trying to identify women by their surnames: most women adopted their husband's surname after marriage, and the Censuses do not report women maiden names. In order to reduce measurement error due to surname changes after marriage, and to avoid excluding married women from the analysis altogether, I match the surname of slaveholders with the surname of the mother of each individual. Although this strategy only deals with recent surname changes, because mass migration only started about 15 years before the census data was collected it is likely that a substantial proportion of the individuals would have encountered their spouse close to their place of origin. Appendix A provides a detailed description of how I construct the data set.

Information about individuals' surnames is available until the census of 1930. I carry out the analysis of migrants in 1930 for two reasons. Firstly, until the 1920 census, migration to the North was very limited; secondly, the microdata available for the 1930 census covers 5% of the individuals, whereas the 1920 microdata only covers 1% of the sample, which does not provide sufficient observations for the analysis. To better capture the effect of gender roles, I focus on individuals living in urban areas, so I only consider locations where labour market conditions would differ from those of the agricultural South. To further ensure that the surname matching captures individuals from cotton and tobacco regions I only include migrants that are registered in non-slave states and whose mother was born in a slave state.

I restrict the sample to African American migrants of age between 25 and 54 and estimate the following equation:

$$\begin{aligned} y_{i,c} = \gamma_0 + \gamma_1 cotton\, migrant \times female_i + \gamma_2 tobacco\, migrant \times female_i + \\ \gamma_3 cotton\, migrant_i + \gamma_4 tobacco\, migrant_i + \gamma_5 female_i + \gamma_6 \mathbf{X}_i + \xi_c + \varepsilon_{i,c} \end{aligned} \tag{3}$$

[3] I will consider as slave states those states where slavery persisted until emancipation: Missouri, Kentucky, Virginia, Maryland, North Carolina, Tennessee, Arkansas, South Carolina, Georgia, Alabama, Mississippi, and Louisiana.

where *cotton migrant* and *tobacco migrant* are dummy variables taking value 1 if the surname of the mother of the individual matches the surname of a cotton or tobacco slaveholder respectively; $\boldsymbol{X}$ is a vector of individual controls (as in equation (1)); and ξ represents county fixed effects.

It is clear that measuring migration by surname matching is subject to measurement error, the extent of which is very hard to estimate. In fact, not all ex-slaves decided to take the surnames of slaveholders, as evidenced by the organizations which arose after emancipation encouraging African Americans to choose different surnames (Van Deburg, 1997)[4]. In my data set I match around 75% of the individuals. Clearly, not all of the matched individuals adopted the surname of their most recent slaveholder: I am therefore capturing an increase in the probability that the matched individual is a descendant of an ex-slave who worked in specific crops. As an indirect test of the extent to which the surname matching captures the origin of African Americans, I use the number of cotton and tobacco surnames by county in 1880 as a predictor of my measure of cotton prevalence by county, and of female labour force participation. Table A.3 shows a positive and significant correlation with both outcomes, a negative correlation of tobacco surnames with the cotton prevalence measure, and no correlation of other surnames with the ratio of African American women to men in the labour force. The positive correlation between cotton surnames and the share of women in the labour force is illustrated in figure C.8.

The second method follows closely from the main empirical strategy of this paper. For the year 1940, information is available about the county of residence 5 years prior to the census year. This allows for a direct study of migration: I estimate equation (1), where the main variable of interest is relative cotton prevalence in the county of origin of the individual, interacted with *female*. To make the analysis comparable to that of 1930, I only include migrants located in non-slave states whose mother was born in a slave state.

One issue is that migrants may be a selected sample of the population, and hence that the differences observed are due to the different characteristics of migrants compared to non-migrants. As my samples are entirely composed of migrants, this is not a major concern. However, the results may be driven by differences of selection into migration across counties. In fact, differences in local labour market conditions across counties may push different types of individuals to migrate, who may be characterised by different skills.

[4]See section 2 for further details.

Notice that when I analyse migrants I compare men and women who migrated from the same areas; therefore, in order for selection to be driving the results, not only there would need to be differences in selection of migrants across regions with historical prevalence of different crops, but selection would also need to differ by gender within region. I address this issue more formally in section 6.1, by analysing whether migrants from cotton and tobacco regions have different observable characteristics.

5 The effect of gender division of labour on labour market outcomes

In this section I describe how cotton production relative to tobacco affects the labour market outcomes of African American women. I document the short term effect using the 1880 sample, and long term effect using data from 1870 until 2010.

5.1 Short and long term effect

The first set of results come from 1880 census data. Table 3 shows the effect of living in areas where cotton was more important relative to tobacco on labour force participation (columns 1 and 2) and log of occupation income score of African American women (columns 3 and 4), conditional on being in the labour force. All reported results include individual controls and county controls interacted with *female*[5]. State fixed effects are included in columns 1 and 3, and columns 2 and 4 add county fixed effects. The coefficients reported correspond to the effect of a one standard deviation increase in cotton production relative to tobacco in 1840. Both specifications show a positive effect on labour force participation of African American women and on their occupation income score. A one standard deviation increase in share of farmland cultivated with cotton relative to tobacco has an additional positive effect on labour force participation for females compared to males of around 8 percentage points. Notice that the coefficient is very similar when including state or county fixed effects: this suggests that the results are not driven by inherent differences in the labour market conditions of counties with a prevalence of cotton or tobacco.[6] Similarly, women's occupation income scores increase with cotton

[5]Excluding individual or county controls interacted with *female* produces qualitatively similar results. See section 4.1 for a description of the controls included.

[6]Note that with county fixed effects the variable *relative cotton production* is dropped from the regression, as it is measured at the county level, and thus the coefficient is not estimated.

prevalence. The effect is economically significant: a one standard deviation increase in relative cotton production increases the average occupation income score by 15% when including state fixed effects, and 16% with county fixed effects.

Columns 1 and 3, which only include state fixed effects, allow for the estimation of the effect of cotton production for African American men, which is negative for both labour force participation and occupation income score. This implies that the higher the cotton production relative to tobacco, the lower the probability that men will be in the labour force, and the lower their average occupation income score. The negative results for men may indicate some substitutability between the labour of African American men and women.

Table 3: OLS Regressions - 1880

	Labour Force Participation		Log Occupation Income Score	
	(1)	(2)	(3)	(4)
Relative Cotton Production 1840×Female	.084***	.079***	.144***	.155***
	(.009)	(.009)	(.015)	(.015)
Relative Cotton Production 1840	-.021***		-.034***	
	(.006)		(.009)	
Female	-.825***	-.822***	-.593***	-.620***
	(.026)	(.027)	(.038)	(.041)
Individual Controls	Y	Y	Y	Y
County FE	N	Y	N	Y
State FE	Y	Y	Y	Y
County Controls*Female	Y	Y	Y	Y
Counties	747	747	746	746
N	186,563	186,563	134,798	134,798
R-Squared	0.4637	0.4857	0.2056	0.2512

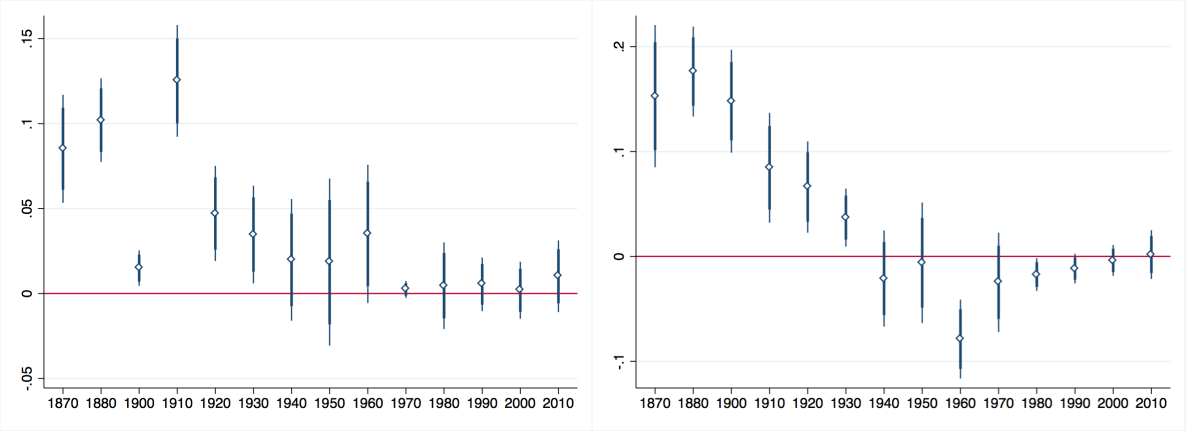
Notes: I include individuals aged between 25 and 54. Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. The coefficients of the variable *relative cotton production* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840. Data from the 1880 Population Census.

Overall, these results indicate that in areas where cotton is more prevalent than tobacco, women are more likely to participate in the labour market, and they are more likely to be employed in occupations that on average pay higher wages. However, it is important to note that because occupation income score measures the median wage paid for each occupation across the country, this analysis does not capture differences in wage rates due to gender or location. Looking at the type of occupations (figure C.5), one sees that in counties with higher cotton prevalence women are more likely to work in agricul-

ture (about 55%), whereas in counties with lower cotton prevalence the most common occupation is service in private households (66%). A much smaller percentage of women (roughly 4%) are employed in other occupations - mostly manufacturing and services - in both regions. Figure C.6 shows the occupations of African American men by cotton prevalence: the patterns are similar to those for women, but the differences in the type of occupation by relative cotton prevalence are not as stark. These graphs suggest that cotton areas are overall more agricultural compared to tobacco. However, the census is missing information for a much higher percentage (12% more) of men in counties where relative cotton prevalence is lower, and hence tobacco prevalence is higher.

The results described so far refer to 1880, 15 years after slavery was abolished in all states. In order to analyse persistence, I run regressions with cross sectional data from all census years between 1870 and 2010. Figure 2 plots the coefficients of the variable $relative\,cotton\,production \times female$ obtained by estimating equation (1) with OLS including individual controls and county fixed effects, where the dependent variables are labour force participation (left panel) and the log of occupation income score (right panel).

FIGURE 2: Effect on African American Women's Labour Market Outcomes 1870-2010



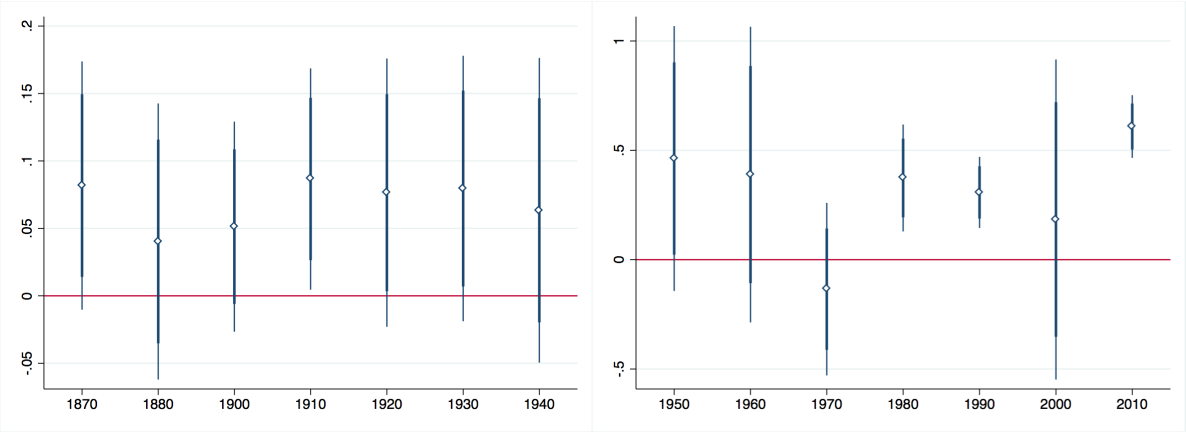
(A) Labour Force Participation (B) Occupation Income Score

Notes: These two graphs show the coefficient plots of regressing labour force participation (left) and log of occupation income score (right) on the variable $relative\,cotton\,production \times female$, produced with OLS regressions using census data for years 1870-2010. I include individuals aged between 25 and 54. The size corresponds to the effect of a one standard deviation change in the dependent variable. The confidence intervals reported are 95 and 99%. I include individual controls and county fixed effects. Standard errors are clustered at the county level.

Both graphs exhibit a similar pattern: the positive effect of cotton for females on both labour force participation and occupation income score is stronger than the effect

for males, and it is particularly evident up to the beginning of the 20th century, when it starts to phase out. The occupation income score effect becomes negative in 1960 and 1970, only to phase towards zero in the following decades. Persistence is stronger for labour force participation: the effect is positive and significant until 1930, and it remains positive, although noisier [7], until 1960, after which it starts approaching zero. The effect on occupation income score is positive until 1930 and reversed in 1960: it is not statistically different from zero in later years. As robustness check, I include in the regression state fixed effects interacted with the *female* dummy variable. Figure C.9 shows similar patterns, although the coefficient is not always statistically significant from zero. This is because the fixed effects capture the variation in cotton versus tobacco production across states.

FIGURE 3: Effect on Female to Male Workers Ratio Within Occupation



(A) 1870-1940 (B) 1950-2010

Notes: These two graphs show the coefficient plots of regressing the share of female to male labour in each occupation on the variable $relative\,cotton\,production \times female$, produced with OLS regressions using census data for years 1870-2010. I include individuals aged between 25 and 54. The size corresponds to the effect of a one standard deviation change in the dependent variable. The confidence intervals reported are 95 and 99%. I include state and occupation fixed effects. Standard errors are clustered at the county level.

To further investigate whether the effect on labour force participation is driven by differences in occupations across counties, or by women participating more in the labour force within occupations, I analyse the effect of relative cotton to tobacco production on the share of female to male workers within occupations by county. I regress the

[7] Part of the reason for the large standard errors after 1940 is that after that year the number of counties which are identified in the population censuses decreases dramatically.

ratio of female to male workers in each occupation-county combination on relative cotton production and state and occupation fixed effects, hence I exploit variation across county within each state. Figure 3 shows that the effect is consisently positive until 1960, and in almost all years it is significant at the 5% level. Later years' estimates are positive (except for 1970), although noisier, due to the limited number of counties identified in the samples after 1940.

While discussing persistence, it is important to understand the possible role of migration. Firstly, migration from the South to the North may have contributed to the phasing out of the effects. Since the 1910s and up until the 1970s, the Southern states of the U.S., and especially rural areas, have been the subject of mass out-migration of African Americans, who settled in urban areas of the Midwest, Northeast and West (the Great Migration). In fact, I find higher migration rates from counties with higher cotton prevalence compared to tobacco. Secondly, migration within the South East cannot be excluded, even at the end of the 19th century. In fact, minor migration waves appear to have happened around 1880 from the South to Kansas, Oklahoma and Colorado (Johnson and Campbell, 1981). However, data from the 1900 census suggests that by that year 90% of African Americans still lived in the South. Finally, note that the high migration rates in cotton regions imply that the population in later decades may be a selected sample of the original population.

5.2 Comparing African American and White women

In order to determine whether the effect is specific to African American women, I investigate how labour market outcomes of African American women differ from labour market outcomes of white women in cotton counties. Even though labour force participation of white women is very low until the mid of the 20th century (below 30% until the 1950s) (Boustan and Collins, 2014), a small percentage is reported to work in agriculture.

I run regressions including black and white women where I interact the measure of relative cotton prevalence with a dummy variable taking value 1 if the woman is African American. Table 4 shows that African American women in cotton regions are more likely to participate in the labour force, compared to white women, by about 6.7 percentage points (column 1), and their occupation income score is about 12% higher (column 3). Following Kleven and Landais (2016), I investigate whether there are significant differences between African American and white women in the number of children. Contrary

Table 4: Women: African American and white - 1880

	Labour Force Participation		Log Occupation Income Score		N Children	
	(1)	(2)	(3)	(4)	(5)	(6)
Relative Cotton Production 1840 × Black	.067***		.119***		.117***	
	(.007)		(.014)		(.020)	
Relative Cotton Production 1840		.037***		.029***		.003*
		(.009)		(.017)		(.027)
Black	.059***		-.416***		-.146***	
	(.020)		(.038)		(.056)	
Individual Controls	Y	Y	Y	Y	Y	Y
County FE	Y	N	Y	N	Y	N
State FE	Y	Y	Y	Y	Y	Y
County Controls×*black*	Y	Y	Y	Y	Y	Y
Counties	779	735	777	647	779	737
N	300,841	93,884	65,243	42,625	342,879	102,473
R-Squared	0.4217	0.2407	0.1625	0.0782	0.2168	0.1430

Notes: I include individuals aged between 25 and 54. Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. The coefficients of the variable *relative cotton production* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

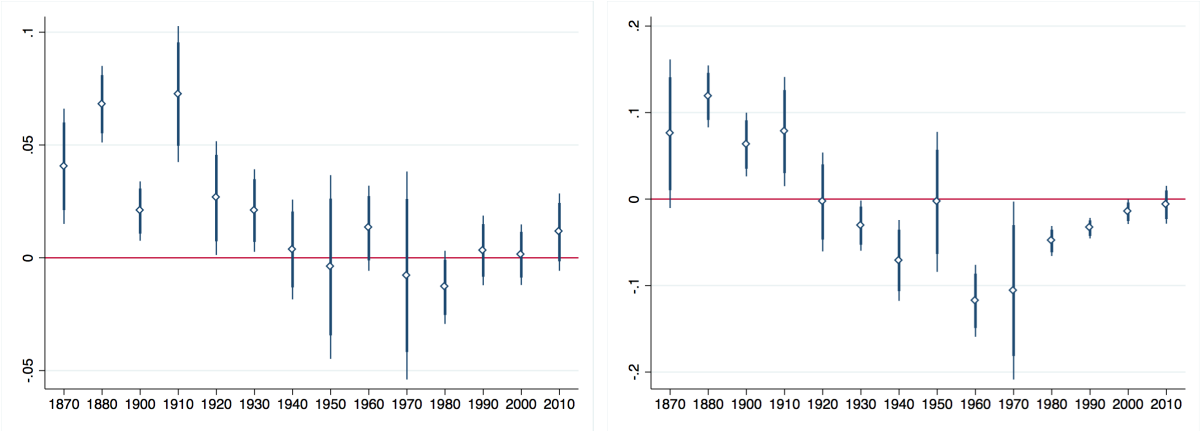
to their findings, here I observe a positive differential effect on the number of children of African Americans. Furthermore, I perform a similar exercise, but comparing African American women in cotton counties to other African American women: the main explanatory variable of interest is now relative cotton production at the county level. Columns 2 and 4 of Table 4 show that there are positive effects on labour force participation and occupation income score, but no effect on the number of children[8].

I then analyse how the labour market responses of African American women compare to those of white women in cotton regions over time. Figure 4 shows the coefficient plot of cross-sectional regressions where the variable of interest is the interaction of *relative cotton production* and the dummy variable *black*. The differential effect is similar to that between African American women and men. The left panel shows that labour force participation is again higher for African American women up to 1940[9], until it phases out in the following decades, with the exception of 1970 and 2010. The differential effect on occupation income score for African Americans is also positive and significant at the beginning of the period analysed, and it reaches the maximum size of about 10% in 1880. However, it becomes negative in the mid-20th century, and approaches zero at the end of the century.

[8]The results columns 2, 4 and 6 of Table 4 are produced by controlling for state, and not county, fixed effects, as the measure of relative cotton prevalence would be collinear with county fixed effects.

[9]The coefficient is significantly different from zero at the 10% significance level.

Figure 4: Effect on African American Women Compared to White Women



(a) Labour Force Participation (b) Occupation Income Score

Notes: These two graphs show the coefficient plots of regressing labour force participation (left) and log of occupation income score (right) on the variable *relative cotton production* × *black*, produced with OLS regressions using census data for years 1870-2010. I include individuals aged between 25 and 54. The size corresponds to the effect of a one standard deviation change in the dependent variable. The confidence intervals reported are 95 and 99%. I include individual controls and county fixed effects. Standard errors are clustered at the county level.

5.3 Robustness checks

In order to reduce endogeneity concerns for the OLS estimates caused by omitted variable bias or measurement error, I estimate instrumental variable regressions where I use cotton suitability relative to tobacco as an instrument for relative production.

Table 5 shows the two-stage least squares (2SLS) results. Panel A illustrates the results of the second stage, while panel B shows the first stage. The first stage results suggest that the measure of relative cotton suitability is a strong predictor of the relative cotton shares of farmland: a one standard deviation increase in relative cotton suitability increases relative cotton prevalence by .28-.32 standard deviations, depending on the specification. The correlation between the two measures is strong: the F-test of excluded instruments is always well above the critical value of 10 suggested by Stock et al. (2002). The 2SLS estimates are always positive and significant, and larger than the OLS, both for labour force participation and for occupation income score. This is consistent with the OLS estimates being biased towards zero due to measurement error of cotton production. The coefficients are larger when including county fixed effects (columns 2 and 4). Figure C.10 shows the persistence of the effect estimated with IV regressions. The pattern is very similar to that of figure 2, except that the negative result on occupation income score in

1960 is not statistically significant.

Table 5: Instrumental Variables Regressions - 1880

	Labour Force Participation		Log Occupation Income Score	
	(1)	(2)	(3)	(4)
	Panel A: Two-Stage Least Squares			
Relative Cotton Production 1840×Female	.175***	.177***	.188***	.259***
	(.035)	(.039)	(.050)	(.053)
	Panel B: First Stage			
Relative Cotton Suitability×Female	.328***	.286***	.264***	.259***
	(.039)	(.041)	(.042)	(.053)
T-test of Excluded Instruments	36.76	49.03	23.39	33.86
Individual Controls	Y	Y	Y	Y
County FE	N	Y	N	Y
State FE	Y	Y	Y	Y
County Controls*Female	Y	Y	Y	Y
Counties	747	747	746	746
N	186,563	186,563	134,798	134,798

Notes: I include individuals aged between 25 and 54. Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. The coefficients of the variables *relative cotton production* and *relative cotton suitability* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

I then perform robustness checks using other measures of cotton prevalence. Table B.2 shows the results. Columns 1 and 4 show the estimates when using the measure based on the share of farmland cultivated with cotton and tobacco in 1880; columns 2 and 5 summarize the reduced form estimates, using relative crop suitability, and columns 3 and 6 report the results using the measure of relative cotton prevalence as a share of total production[10]. All three measures produce qualitatively similar results.

To further test for the robustness of the labour force participation results, I estimate a probit model with labour force participation as a dependent variable. Column 1 of Table B.1 shows the results: the coefficient is positive and statistically significant, although smaller in magnitude (3.3 percentage points). Furthermore, one concern is that the occupation income scores results could be driven by a small number of occupations with very high median wages. I replace the occupation income score measure with a variable which ranks each occupation based on its median wage, instead of taking the value of the median wage. Column 4 of Table B.1 shows a similar effect when using the occupational ranking measure: the increase is large and it amounts to approximately 20 percent.

[10]The measures are described in sections 3.1 and 4.2.

One may be concerned that the results could be driven by African Americans migrating to areas offering better job prospects after emancipation. To address this concern, I restrict the analysis to slave states in 1850 and 1860. Figure C.9 shows a map of the slave population in 1860 (top panel) and the African American population in 1880 (bottom panel) by county. It is evident from the two maps that slave population in 1860 predicts African American population in 1880 (major migration waves only started in 1910). However, some counties in the West show a very small percentage of African Americans in 1880 even if no slavery was present at the time, suggesting that, although limited, some migration took place between emancipation and 1880. Column 2 and 4 of Table B.1 show a very similar coefficient for labour force participation compared to the baseline OLS, and a slightly smaller coefficient, but statistically and economically significant, for occupation income score.

An additional issue is that the positive effect of cotton on female labour force participation may be partly or entirely driven by income: women might participate more in the labour market when family income is very low. As no measure of income at the individual level is available for the late 19th century censuses, I run regressions including married African American women and controlling for husband's occupation income score as a proxy for family income.[11] The results are summarized in columns 3 and 5 of Table B.1. The coefficients are positive and statistically significant, suggesting that spouse's income is not a main driver of the results.[12] Interestingly, spouse income has a negative effect on women's probability of participating in the labour market, but a positive effect on occupation income score. However, husband's income could be considered a "bad control": in fact, it is likely to be correlated with individuals' own labour market outcomes and could therefore be an outcome itself.

I then examine whether the results might be driven by differences in education. To test this hypothesis I estimate equation (1) with literacy (1880-1930) or years of education (1940-2010). Figure C.12 shows the coefficient plots. Notice positive and significant coefficients for literacy and education in some of the years (1870, 1940, 1960-2000); however, the positive results are not consistent throughout the sample, showing no definitive evidence of differences in formal education driving the results.

I also test whether cotton prevalence has an effect on wages. The bottom panels

[11] Note that spouse income is only available for currently married individuals.

[12] Note that the sample size is significantly reduced when restricting the regression to married African American women.

of Figure C.12 show no consistently positive effects on wages: the effects are similar for wages across and within occupations. This is consistent with the occupation income scores results, as the positive effect is persistent until 1930.

6 Mechanisms

In this section I explore the mechanisms that drive the effect of gender division of labour on women's labour market outcomes and its persistence over time.

Firstly, to disentangle gender roles from local labour demand effects, I show the results for the two samples of migrants in 1930 and 1940.

Secondly, I investigate whether women in cotton counties have higher labour market experience compared to their counterparts in tobacco counties, resulting in jobs paying higher median wages. To test for this channel, I analyse female labour force participation for young women across counties with different levels of relative cotton production and heterogenoeus effects of age on occupation income scores.

Thirdly, I explore whether the results can be explained with intergenerational transmission mechanisms, by analysing whether labour market status of mothers affects their daughters' labour market outcomes, and whether the effects are heterogeneous in cotton and tobacco regions.

Furthermore, I investigate the possibility of differences in demand for labour across counties, in particular for agriculture.

Finally, I analyse two alternative mechanisms: access to social networks and discrimination in favour African American women.

6.1 Migrants

Analysing the labour market outcomes of migrants from cotton and tobacco areas is key to understanding whether the results described in the previous sections are driven by differences in gender roles in cotton and tobacco regions, or whether they are a consequence of differences in labour demand and labour market conditions across counties.

Table 6 shows the results. Columns 1 and 2 shows a positive relation between having a mother with a surname which matches the surname of a cotton slaveholder and participating in the labour market. Note that the positive relation only appears to be significant for female migrants, and not for male migrants: this suggests that women with mothers

Table 6: Effect on Migrants Living in Urban Areas, 1930 and 1940

	Labour Force Participation	Log Occupation Income Score
	(1)	(2)
	Panel A: 1930	
Mother with Cotton Surname×Female	.050***	.119***
	(.016)	(.027)
Mother with Tobacco Surname×Female	-.011	.075
	(.053)	(.087)
Mother with Cotton Surname	-.004	.016
	(.009)	(.010)
Mother with Tobacco Surname	-.013	.054*
	(.021)	(.029)
Female	-.015	-.809***
	(.020)	(.052)
Individual Controls	Y	Y
State FE	Y	Y
N	85,573	64,622
R-Squared	0.3041	0.5265
	Panel B: 1940	
Relative Cotton Prod.×Female	.010**	-.016***
	(.005)	(.006)
Relative Cotton Prod.	-.006*	.003
	(.003)	(.004)
Female	-.110***	-.588***
	(.026)	(.039)
Individual Controls	Y	Y
County FE	Y	Y
N	54,893	43,920
R-Squared	0.3635	0.4946

Notes: The variable *relative cotton share* measures the relative prevalence of cotton vs. tobacco in the county of origin of the individual. I include state (1930) and county of destination (1940) fixed effects. The sample is restricted to African Americans of age 25-54 who migrated from a slave state to urban areas in non-slave states, whose mother was born in a slave state, and whose information about mother labour force status is known. Data from the U.S. 1930 and 1940 Population Census. Standard errors clustered at the state (1930) and county (1940) level.***p>0.01 **p>0.05 *p>0.10

who had a surname associated with cotton are significantly more likely to be in the labour force. The size of the coefficient is 5%. The results show no effect for individuals whose mother's surname matches the surname of a tobacco slaveholder. Occupation income score conditional on labour force participation is also positive and significantly higher for females with cotton surnames. This suggests positive effects potentially due to female workers having higher labour market experience compared to other migrants.

Columns 3 and 4 show the results of the analysis of individuals who migrated up to 5 years prior to the census of 1940. I estimate equation (1) defining relative production of

Table 7: Characteristics of Migrants Living in Urban Areas, 1930 and 1940

	Married	Age	Education	N Children
	(1)	(2)	(3)	(4)
		Panel A: 1930		
Mother with Cotton Surname×Female	.032	-.144	.030***	.351***
	(.022)	(.235)	(.011)	(.060)
Mother with Tobacco Surname×Female	.149**	1.735*	.033	.379**
	(.060)	(.980)	(.024)	(.173)
Mother with Cotton Surname	-.292***	-4.364***	.022***	-.606***
	(.013)	(.196)	(.007)	(.059)
Mother with Tobacco Surname	-.370***	-4.842	.010	-.701***
	(.035)	(.671)	(.022)	(.108)
State FE	Y	Y	Y	Y
N	85,573	85,573	85,573	85,573
R-Squared	0.0157	0.0188	0.0519	0.0148
		Panel B: 1940		
Relative Cotton Prod.×Female	-.015***	.206**	-.007	.010
	(.004)	(.071)	(.020)	(.009)
Relative Cotton Prod.	.032***	-.841***	-.291***	.112***
	(.005)	(.082)	(.027)	(.014)
Female	-.035***	-.242	.264***	.102***
	(.010)	(.500)	(.041)	(.018)
County FE	N	N	Y	Y
N	54,893	54,893	54,893	54,893
R-Squared	0.0504	0.0606	0.1330	0.0673

Notes: The variable *relative cotton production* measures the relative prevalence of cotton vs. tobacco in the county of origin of the individual. I include individuals aged between 25 and 54. The regressions include county of destination fixed effects. The sample is restricted to African Americans who migrated from a slave state to urban areas in non-slave states, whose mother was born in a slave state, and whose information about mother labour force status is known. Data from the U.S. 1930 and 1940 Population Census. Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10

the two crops according to the values in the county of origin. Labour force participation is higher for female migrants from cotton areas: a one standard deviation increase in the index of relative cotton production increases the probability of being in the labour market for women by 1 percent. However, in this sample occupation income score of women originating from cotton counties is lower. The latter possibly reflects the fact that I include in the analysis only recent migrants, whereas the 1930 migrants may be a combination of recent and longer-term migrants.

Although the regressions reported control for individual characteristics, allowing them

have a different effect for women and men, it is important to understand whether there are differences in the selection of migrants originating from cotton and tobacco regions. To do this, I test whether cotton prevalence relative to tobacco predicts observable characteristics of migrants: whether the individual is married, age, literacy (1930) and years of education (1940), and number of children. Table 7 shows differences in number of children and literacy of migrants from cotton regions compared to other migrants. However, the 1940 sample of recent migrants shows somewhat different patterns. In fact, the results suggests that migrants from cotton areas have less formal education on average. Women coming from higher cotton production regions have a lower probability of being married and are on average slightly older, whereas men have higher probability of being married, are younger, have lower education and a higher number of children.

These results suggest differences in gender roles between migrants from cotton and tobacco regions, indicating that the higher female labour force participation found in cotton regions does not solely reflect differences in labour market conditions.

6.2 Labour market experience

In order to understand whether the occupation income score results can be attributed to greater labour market experience of women in cotton counties, I explore the effect of an increase in age, and whether this is different for women in regions with cotton or tobacco prevalence, using data from 1880. The intuition is the following: a larger effect of age on occupation income score in cotton areas could indicate that women work more not only along the extensive margin, but also along the intensive margin. Because the data does not enable me to distinguish between the two margins, I also analyse whether women in cotton counties are more likely to work even at a young age - between 14 and 24 years old.

Table 8 shows that young women in cotton areas are more likely to be in the labour force. Moreover, column 4 shows that the effect of age on occupation income scores is higher for women in cotton counties: they are 1 percent higher for each standard deviation increase in age (9 years). This could indicate a learning-by-doing effect, which may increase women's productivity in cotton areas as females have more experience in the labour market. An alternative interpretation would suggest that as age and labour market experience increase, women have more knowledge about job opportunities, which leads to better matching.

Table 8: Labour Market Experience: Young Women and Heterogeneous Effects for Age 1880

	Labour Force Participation		Log Occupation Income Score	
	Age 15-24	Age 25-54	Age 15-24	Age 25-54
	(1)	(2)	(3)	(4)
Relative Cotton Production 1840×Female	.069***	.080***	.154***	.130***
	(.008)	(.012)	(.014)	(.019)
Relative Cotton Production 1840×Female×Age		-.000		.010**
		(.000)		(.003)
Relative Cotton Production 1840×Age		-.000***		-.000
		(.000)		(.000)
Female×Age		-.063***		-.004***
		(.004)		(.001)
Age		.034***		.023***
		(.002)		(.004)
Female	-.407***	-.588***	-.430***	-.492***
	(.020)	(.031)	(.034)	(.047)
Individual Controls	Y	Y	Y	Y
County FE	Y	Y	Y	Y
County Controls*Female	Y	Y	Y	Y
Counties	737	747	728	746
N	121,065	186,563	98,505	134,798
R-Squared	0.3059	0.4894	0.1773	0.2517

Notes: Standard errors clustered at the county level. ***p>0.01 **p>0.05 *p>0.10. The coefficients of the variable *relative cotton production* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

6.3 Intergenerational transmission

In this section I analyse the intergenerational transmission mechanisms for labour market outcomes, and whether these are different in cotton and tobacco areas. To do this, I look for heterogeneous effects for individuals whose mother is in the labour force by including a triple interaction in the baseline model, using data from 1880. The results are summarized in Table 9. The results show that mother labour force participation has an additional positive effect on occupation income score of daughters in cotton-prevalent counties, but not on labour force participation. The increase in occupation income score is substantial: about 7 percent.

The fact that mother labour force participation matters more in cotton areas might again indicate higher labour market experience of mothers, which can be transmitted to their daughters. Labour market experience may include knowledge of specific jobs, or general knowledge of the labour market, which translates into higher occupation income score for their daughters.

Table 9: Heterogeneous Effects for Mother Labour Force Status

	Labour Force Participation	Occupation Income Score
	(1)	(2)
Relative Cotton Production 1840×Female	.027***	.106***
	(.004)	(.014)
Relative Cotton Production 1840×Female×Mother LFP	.001	.067***
	(.005)	(.018)
Relative Cotton Production 1840×Mother LFP	-.001	.002
	(.004)	(.008)
Mother LFP×Female	.027**	-.163***
	(.013)	(.021)
Mother LFP	-.001	.066***
	(.005)	(.018)
Female	-.132***	-.295***
	(.010)	(.034)
Individual Controls	Y	Y
County FE	Y	Y
Counties	720	684
N	115,528	76,509
R-Squared	0.5415	0.1596

Notes: Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. The coefficients of the variable *relative cotton production* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. LFP indicates labour force participation. Data from the 1880 Population Census.

Another possible interpretation regards access to networks: being in the labour force in cotton areas gives women better access to social networks to find better jobs for their daughters, which translate into higher occupation income scores. The differential effect may be a result of a denser network in cotton areas, due to the presence of more women in the labour force. Interestingly, mother labour force status has a negative effect on occupation income score of daughters in counties with no cotton production.

The increase in labour force participation can partly be explained by mother labour force status: in fact, the size of the coefficient of cotton prevalence interacted with female is reduced compared to the baseline results, and one can observe a positive and significant effect of the interaction of mother labour force participation with female. This suggests a gender role transmission effect from mother to daughter regarding the decision to participate in the labour market. While I find no significant difference in the transmission of gender roles for areas with higer or lower cotton and tobacco prevalence, intergenerational transmission mechanisms may still explain the persistence of results, due to the higher initial female labour force participation rates in cotton areas.

6.4 Demand for labour

In this section I explore whether the main results could be explained by higher demand for labour in cotton counties. Although analysing the effects for migrants makes it possible to disentangle demand and supply mechanisms, labour demand, in particular for agriculture, may still partly explain why African Americans have better labour market outcomes in cotton areas. Women may have higher labour force participation rates and occupation income scores due to the prevalence of agriculture in the South in the first half of the 20th century.

Firstly, differences in demand for labour are likely to exist between areas with high and low relative cotton prevalence. I indirectly test this channel by firstly analysing the effect for African American men, and secondly whether women of other ethnicities display similar labour market outcomes to African American women.

I analyse heterogeneous effects on occupation income score for women working in agriculture. Table B.3 shows the results of the regressions including the triple interaction of *relative cotton production*, *female* and a dummy variable indicating whether the individual is employed in agriculture. The results indicate that the increase in occupation income score is not driven by individuals working in agriculture: in fact, the coefficient of the main variable of interest remains positive and significant. This suggests that the results are not only driven by a mechanical effect deriving from higher labour demand for women in agriculture in cotton regions.

I investigate the effects on women of other ethnicities using the 1880 census data to estimate equation (1), but restricting the sample to exclude African Americans. The remaining ethnicities are white, Native American, Chinese and Japanese. Table B.4 shows no correlation between female labour force participation and relative cotton prevalence, but a positive and significant effect on occupation income score. I explore this further by including in the model a triple interaction of the main explanatory variable of interest with a dummy variable indicating that the individual works in the agricultural sector. Column 3 shows that the entire increase in occupation income score is due to women working in agriculture, who on average have higher occupation income scores than men. Moreover, the triple interaction has a negative coefficient, which indicates that women working in agriculture where cotton is relatively more prevalent than tobacco have a lower occupation income score. In conclusion, relative cotton prevalence seems to affect African American women differently compared to women of other ethnicities, which indicates that

the results are not only driven by higher labour demand for agriculture.

6.5 Discrimination and access to social networks

In this section I explore two other channels that may explain the positive effect of cotton on black women's labour market outcomes: lower discrimination towards African American women and better access to social networks.

Discrimination may be lower where cotton was more prevalent because cotton regions were characterized by a higher concentration of female slaves; hence it is likely that African American women had more opportunities to interact with white women and men. In turn, this could result in lower discrimination, which may imply a higher probability of being hired by white employers. For an indirect test of discrimination, I investigate whether relative cotton production predicts the ratio of mixed couples composed of white husband and African American wife by county in 1880. In this regression I control for state fixed effects, to reduce the possibility of capturing effects of laws or customs regarding mixed marriages. Column (1) of Table B.5 shows a positive and significant effect, which corresponds to a 30% increase in the ratio of mixed couples in counties with higher cotton prevalence. Columns 2 and 3 show the results of regressing labour force participation and occupation income score on the measure of relative cotton prevalence times *female* and the ratio of mixed couples interacted with *female*. Both measures have a positive effect on both outcomes, and the size of the interaction of relative cotton production and *female* is very similar in size to the baseline results, suggesting that this measure of discrimination cannot entirely explain the increase in labour force participation and occupation income score. Note that these results merely suggest differences in discrimination. In fact, a higher ratio of mixed couples in cotton areas may simily be due to the higher number of African American women.

Moreover, because cotton areas had a higher concentration of African Americans women, they may have benefitted from better access to social networks. Chay and Munshi (2013) put forward a similar hypothesis: they find evidence of African Americans being able to better exploit social networks in areas where the crops grown were more labour intensive, due to the higher concentration of African American individuals. To test the social network channel I estimate equation (1) adding as explanatory variable the number of female slaves in a county in 1860 interacted with the female dummy. Columns (2) and (3) show that controlling for cotton prevalence, the number of female slaves is positively

associated with both labour force participation and occupation income score. This may indicate that better access to social networks plays a role. However, the coefficient of cotton prevalence times female decreases in magnitude, but remains significant, signalling that although the two measures are correlated, cotton prevalence has a positive effect on labour market outcomes which is independent of the higher density of women.

7 Conclusion

This paper stresses the importance of culture and highlights the role of working practices in the formation of gender roles. The findings suggest strong path dependence in regard to labour market outcomes.

By analysing the case of African American women in the U.S., this paper contributes to the literature on the origin and the persistence of gender roles in women's labour market outcomes. I find higher labour force participation and higher occupation income scores of African American women after slavery in areas where gender division of labour was less well-defined during slavery, which I identify as areas producing cotton. I find persistent effects until at least 70 years after the end of slavery for occupation income scores, and at least 90 years for labour force participation. Due to the lack of migration data for the first decades of the 20th century, I construct a data set of migrants by matching surnames of African Americans in urban areas of free states in 1930 (during the Great Migration) with surnames of slaveholders in the 1860 census. In addition, I analyse direct data on migration from the 1940 census. I find that migrants who relocated to urban areas from counties with high cotton prevalence are more likely to participate in the labour force compared to migrants from other regions, which supports the hypothesis that gender roles might be driving the results.

My findings are consistent with a model in which more widespread cotton cultivation affects gender roles, which in turn affects the labour market participation decison of women. Consequently, due to a learning-by-doing effect which increases women's productivity, women who work more are able to obtain jobs paying higher wages. Moreover, I find evidence to suggest that mothers' labour market experience affects their daughters' productivity, resulting in higher occupation income scores.

Although the mechanisms summarized above are the focus of this paper, the patterns observed in the data indicate that other channels may have also affected the labour market outcomes of African American women, in addition to gender roles. I consider differences in

discrimination and in access to social networks, and, although I find evidence to suggest that both channels playing a role, they do not appear to be sufficient to explain the patterns outlined in the paper, in particular the labour market outcomes of migrants.

References

Ager, P., Brueckner, M. and Herz, B. (2016), 'The boll weevil plague and its effect on the Southern agricultural sector, 1889–1929', *Explorations in Economic History* .

Alesina, A., Giuliano, P. and Nunn, N. (2013), 'On the origins of gender roles: Women and the plough', *The Quarterly Journal of Economics* **128**(2), 469–530.

Bertocchi, G. (2015), 'The legacies of slavery in and out of Africa', IZA Discussion Paper.

Biavaschi, C., Giulietti, C. and Siddique, Z. (2013), 'The economic payoff of name Americanization', IZA Discussion Paper.

Blake, K. L. (1996), *First in the Path of the Firemen: The Fate of the 1890 Population Census*, Washington, DC: National Archives and Records Administration.

Boustan, L. P. and Collins, W. J. (2014), The origin and persistence of black-white differences in women's labor force participation, *in* 'Human Capital in History: The American Record', University of Chicago Press, pp. 205–240.

Chay, K. and Munshi, K. (2013), 'Black networks after emancipation: Evidence from reconstruction and the great migration', *Unpublished working paper*.

Christian, C. (2014), 'Lynchings, labour, and cotton in the US South', mimeo.

Collins, W. J. and Wanamaker, M. H. (2014), 'Selection and economic gains in the great migration of African Americans: New evidence from linked census data', *American Economic Journal: Applied Economics* **6**(1), 220–252.

Farré, L. and Vella, F. (2013), 'The intergenerational transmission of gender role attitudes and its implications for female labour force participation', *Economica* **80**(318), 219–247.

Fenske, J., Kala, N. et al. (2014), 1807: Economic shocks, conflict and the slave trade, Technical report.

Fernández, R. (2013), 'Cultural change as learning: The evolution of female labor force participation over a century', *The American Economic Review* **103**(1), 472–500.

Goldin, C. (1977), 'Female labor force participation: The origin of black and white differences, 1870 and 1880', *The Journal of Economic History* **37**(01), 87–108.

Goldin, C. (1986), The female labor force and american economic growth, 1890-1980, *in* 'Long-term factors in American economic growth', University of Chicago Press, pp. 557–604.

Goldin, C. (1990), *Understanding the gender gap: An economic history of American women*, New York: Cambridge University Press.

Goldin, C. D. (1991), 'The role of World War II in the rise of women's employment', *The American Economic Review* pp. 741–756.

Goldin, C. and Sokoloff, K. (1984), 'The relative productivity hypothesis of industrialization: The American case, 1820 to 1850', *The Quarterly Journal of Economics* **99**(3), 461–487.

Gutman, H. G. et al. (1976), *Black family in slavery and freedom, 1750-1925*, Pantheon Books.

Higgs, R. (1976), 'The boll weevil, the cotton economy, and black migration 1910-1930', *Agricultural History* **50**(2), 335–350.

Johnson, D. M. and Campbell, R. R. (1981), 'Black migration in America: A social demographic history', *Studies in Social and Economic Demography* (4).

Jones, J. (2009), *Labor of love, labor of sorrow: Black women, work, and the family, from slavery to the present*, Basic Books.

Kleven, H. J. and Landais, C. (2016), 'Children and gender inequality: Evidence from Denmark', *Unpublished working paper*.

Lange, F., Olmstead, A. L. and Rhode, P. W. (2009), 'The impact of the boll weevil, 1892–1932', *The Journal of Economic History* **69**(03), 685–718.

Metzer, J. (1975), 'Rational management, modern business practices, and economies of scale in the ante-bellum southern plantations', *Explorations in Economic History* **12**(2), 123–150.

Nunn, N. (2008), 'The long-term effects of Africa's slave trades', *The Quarterly Journal of Economics* **123**(1), 139–176.

Nunn, N. and Wantchekon, L. (2011), 'The slave trade and the origins of mistrust in Africa', *The American Economic Review* **101**(7), 3221–3252.

Ransom, R. L. and Sutch, R. (2001), *One kind of freedom: The economic consequences of emancipation*, Cambridge University Press.

Ruggles, S., Sobek, M., Alexander, T., Fitch, C. A., Goeken, R., Hall, P. K., King, M. and Ronnander, C. (2010), 'Integrated public use microdata series: Version 3.0. minneapolis: Minnesota population center'.

Shlomowitz, R. (1979), 'The origins of Southern sharecropping', *Agricultural History* **53**(3), 557–575.

Stock, J. H., Wright, J. H. and Yogo, M. (2002), 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business & Economic Statistics* **20**(4), 518–530.

Van Deburg, W. L. (1997), *Modern Black Nationalism: From Marcus Garvey to Louis Farrakhan*, number 2, NYU Press.

Walsh, L. S. (1989), 'Plantation management in the Chesapeake, 1620–1820', *The Journal of Economic History* **49**(02), 393–406.

Walsh, L. S. (2001), 'The Chesapeake slave trade: Regional patterns, African origins, and some implications', *The William and Mary Quarterly* **58**(1), 139–170.

Wayne, T. K. (2007), *Women's roles in nineteenth-century America*, Greenwood Publishing Group.

A Data appendix

A.1 Labour force participation variable

The outcome variable *labour force participation* is constructed using micro data from IPUMS for the years 1870-2010. Note that no micro data is available for year 1890, because a large part of the census forms were lost in a fire; therefore, that year is missing from the analysis (Blake, 1996).

The definition of labour force participation is different for censuses before and after 1940. Late 19th and early 20th century censuses define labour force participation as individuals engaging in any gainful occupation. From 1940, labour force participation is instead defined as "within a specific reference week, having a job from which one is temporarily absent (e.g., on vacation), working, or seeking work".

The gainful employment definition prior to 1940 may create some bias in establishing the labour force participation of married women. This is because some might have considered their primary occupation "housewife", when in fact they might have been engaged in other type of work. This may produce biased results in my analysis, if one beliefs that differences in gender roles between cotton and tobacco regions may result in higher

Table A.1: Married and Unmarried Women, 1880

	Labour Force Participation		Log Occupation Income Score	
	Married	Unmarried	Married	Unmarried
	(1)	(2)	(3)	(4)
Relative Cotton Production 1840×female	.120***	.058***	.191***	.159***
	(.010)	(.009)	(.021)	(.015)
Female	-.983***	-.325***	-.679***	-.571***
	(.028)	(.024)	(.049)	(.039)
Individual Controls	Y	Y	Y	Y
County FE	N	Y	N	Y
State FE	N	Y	Y	Y
County Controls*Female	Y	Y	Y	Y
Counties	733	720	731	709
N	144,557	40,333	98,808	34,620
R-Squared	0.5792	0.1342	0.2521	0.2373

Notes: Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. I include individuals of age between 25 and 54. Columns 3 and 5 only include African American women. The coefficients of the variables *relative cotton share* and *relative cotton suitability* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

underreporting in tobacco areas. Table A.1 shows the effect of cotton production relative to tobacco splitting the sample between married individuals and unmarried for year 1880. Although there are differences in magnitudes, with married women showing larger effects, unmarried women are also more likely to participate in the labour force, and to have higher occupation income scores conditional on labour force participation, in cotton regions.

Goldin (1986) reports that women in cotton farms in the late 19th century censuses were likely to underreport labour in family farms; in this case, the estimates in this paper for female labour force participation provide a lower bound for the differences between cotton and tobacco regions. Because who designed the 1910 census was aware of this problem, the enumerators were given instructions which implied that women who worked in farms would be considered as in the labour force, even if they did not receive any wage; this means that the 1910 census is likely to over-estimate the number of unpaid female farm workers (Goldin, 1986). In order to correct for this, I exclude all workers in agriculture for year 1910; the results are shown in table A.2. The size of the coefficient is smaller than when analysing the whole sample, but there is still a small effect on labour

Table A.2: Excluding farm workers, 1910

	Labour Force Participation	
	(1)	(2)
Relative Cotton Production 1840×female	.094***	.020*
	(.009)	(.012)
Female	-.541***	-.452***
	(.028)	(.034)
Individual Controls	Y	Y
County FE	N	Y
State FE	N	Y
County Controls*Female	Y	Y
Counties	626	614
N	31,001	16,717
R-Squared	0.2484	0.3168

Notes: Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. I include individuals of age between 25 and 54. Columns 3 and 5 only include African American women. The coefficients of the variables *relative cotton share* and *relative cotton suitability* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

force participation, although now only significant at the 10% level.

A.2 Construction of the 1930 data set of migrants

To overcome data limitations about migration in the early 20th century, I match surnames of cotton and tobacco slaveholders in order to construct a data set of migrants in year 1930. In this section I outline in detail the procedure I use to construct the data set.

I use data from the 1860 slave schedules to obtain slaveholders' surnames and I associate a crop to each surname based on the production of crops in each county in 1860. To achieve this, I determine the most common crop in each county in terms of crop production, and match this crop to the surname of each slaveholder registered in that county. However, in order to determine the origin of individuals whose surname match that of a slaveholder, I need to obtain a data set where each surname is associated to only one crop. Therefore, for those surnames that appear in more than one county, I only consider the slaveholder owning the largest number of slaves.

To decrease the extent of masurement error, I drop the most common surnames from the data set, and those which are associated with famous persons at the time, such as the surnames of current and former Presidents. Finally, I match African Americans whose mother share a surname with a slaveholder, and I identify as "migrants" those individuals who are not registered in a slave state, but whose mother was born in a slave state.

Table A.3: Surnames and Cotton Prevalence

	Relative Cotton Production 1840	Females/Males in the Labour Force
	(1)	(2)
Number of Cotton Surnames	.079***	.011***
	(.009)	(.004)
Number of Tobacco Surnames	-.095***	-.003
	(.021)	(.003)
Number of Wheat Surnames	-.006	-.000
	(.007)	(.001)
Number of Rice Surnames	-.031**	-.001
	(.013)	(.000)
Number of Indian Corn Surnames	-.043	-.011
	(.039)	(.008)
State FE	Y	Y
N Counties	804	931

Notes: The dependent variable in column 2 is the ratio of African American females to males in the labour force Standard errors clustered at the state level.***p>0.01 **p>0.05 *p>0.10

Note that the slave schedules may report more than one slaveholder for some of the slaves; in the case of multiple slaveholders, I only consider the surname of the first slaveholder. Table A.3 shows the correlation between the number African Americans in 1880 whose surname matches that of an cotton plantation owner and cotton production in 1840, and the ratio of females to males in the labour force by county.

B Tables

Table B.1: Probit, Restricting to Slave States and Controling for Income

	Labour Force Participation			Log Occupation Income Score		
	Probit	Slave States	Control for Income	Occupational Ranking	Slave States	Control for Income
	(1)	(2)	(3)	(4)	(5)	(6)
Relative Cotton Production 1840×female	.033**	.083***		1.937***	.106***	
	(.013)	(.008)		(.191)	(.013)	
Relative Cotton Production 1840			.060***			.048*
			(.012)			(.025)
Spouse OIS			-.039***			.258***
			(.008)			(.018)
Female	-.771	-.843***		-6.549***	-.541***	
	(.058)	(.024)		(.485)	(.037)	
Individual Controls	Y	Y	Y	Y	Y	Y
County FE	N	Y	N	Y	Y	N
State FE	N	Y	Y	Y	Y	Y
County Controls*Female	Y	Y	Y	Y	Y	Y
Counties	747	708	1,559	746	584	461
N	186,563	58,460	88,408	134,798	127,972	18,457
R-Squared	0.4731	0.1342	0.1540	0.2219	0.2521	0.2480

Notes: Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. I include individuals of age between 25 and 54. Columns 3 and 5 only include African American women. Spouse OIS refers to the spouse's occupation income score, which is used as a proxy for family income. As county FE cannot be included in columns 3 and 5, state FE and county controls are included. The coefficients of the variables *relative cotton share* and *relative cotton suitability* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

Table B.2: Alternative Measures of Relative Cotton Prevalence

	Labour Force Participation			Log Occupation Income Score		
	(1)	(2)	(3)	(4)	(5)	(6)
Relative Cotton Share 1880×female	.083***			.153***		
	(.008)			(.012)		
Relative Cotton Suitability		.052***			.066***	
		(.009)			(.013)	
Relative Cotton Production ×female			.111***			.134***
As a Share of Tot. Production 1840			(.008)			(.014)
Female	-.805***	-.825***	-1.011***	-.668***	-.584***	-.839***
	(.023)	(.033)	(.032)	(.041)	(.056)	(.067)
Individual Controls	Y	Y	Y	Y	Y	Y
County FE	Y	Y	Y	Y	Y	Y
County Controls*Female	Y	Y	Y	Y	Y	Y
Counties	1,558	1,730	1,021	1,541	1,710	1,015
N	251,559	260,106	204,701	180,365	186,444	147,380
R-Squared	0.4996	0.4974	0.4979	0.2648	0.2597	0.2674

Notes: Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. I include individuals of age between 25 and 54. The coefficients of the variables *relative cotton share* and *relative cotton suitability* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

Table B.3: Occupation Income Score, Heterogeneous Effects for Agriculture

	Log Occupation Income Score
Relative Cotton Share 1880×female	.167***
	(.014)
Relative Cotton Share 1880×female×agriculture	-.201***
	(.014)
Relative Cotton Share 1880×agriculture	.034***
	(.006)
Female×agriculture	.794***
	(.025)
Female	-.900***
	(.030)
Agriculture	-.651***
	(.011)
Individual Controls	Y
County FE	Y
County Controls*Female	Y
Counties	746
N	134,798
R-Squared	0.4846

Notes: Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. I include individuals of age between 25 and 54. The coefficients of the variable *relative cotton production* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

Table B.4: Other Ethnicities

	Labour Force Participation	Log Occupation Income Score	
	(1)	(2)	(3)
Relative Cotton Production 1840 ×female	.005	.062***	-.009
	(.003)	(.011)	(.016)
Relative Cotton Production 1840 ×female×agriculture			-.042***
			(.014)
Relative Cotton Production 1840 ×agriculture			-.034***
			(.003)
Female×agriculture			.708***
			(.024)
Agriculture			-.707***
			(.004)
Female	-.920***	-.349***	-.672***
	(.007)	(.047)	(.050)
Individual Controls	Y	Y	Y
County FE	Y	Y	Y
Y			
County Controls*Female	Y	Y	Y
Counties	828	828	828
N	515,006	273,877	273,877
R-Squared	0.7731	0.2644	0.5863

Notes: Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. I include individuals of age between 25 and 54. The coefficients of the variable *relative cotton production* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

Table B.5: Discrimination

	Ratio of Mixed Race Couples	Labour Force Part.	Log Occ. Income Score
	(1)	(2)	(3)
Relative Cotton Production 1840 × female	-	.079***	.145***
		(.011)	(.015)
Relative Cotton Production 1840	.003*		
	(.001)		
Ratio of Mixed Race Couples 1880 × female		.035***	.081***
		(.008)	(.011)
Female	-	-.869***	-.724***
		(.031)	(.040)
Individual Controls	-	Y	Y
County FE	-	Y	Y
State FE	Y	Y	Y
Industry FE	-	Y	Y
Counties	820	745	744
N	820	184,890	133,428
R-Squared	0.4298	0.4886	0.2569

Notes: Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. Columns 2 and 3 include individuals of age between 25 and 54. The regression estimates in column 1 are at the county level and refer to mixed race couples where the wife is African American and the husband is white. The coefficients of the variable *relative cotton production* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

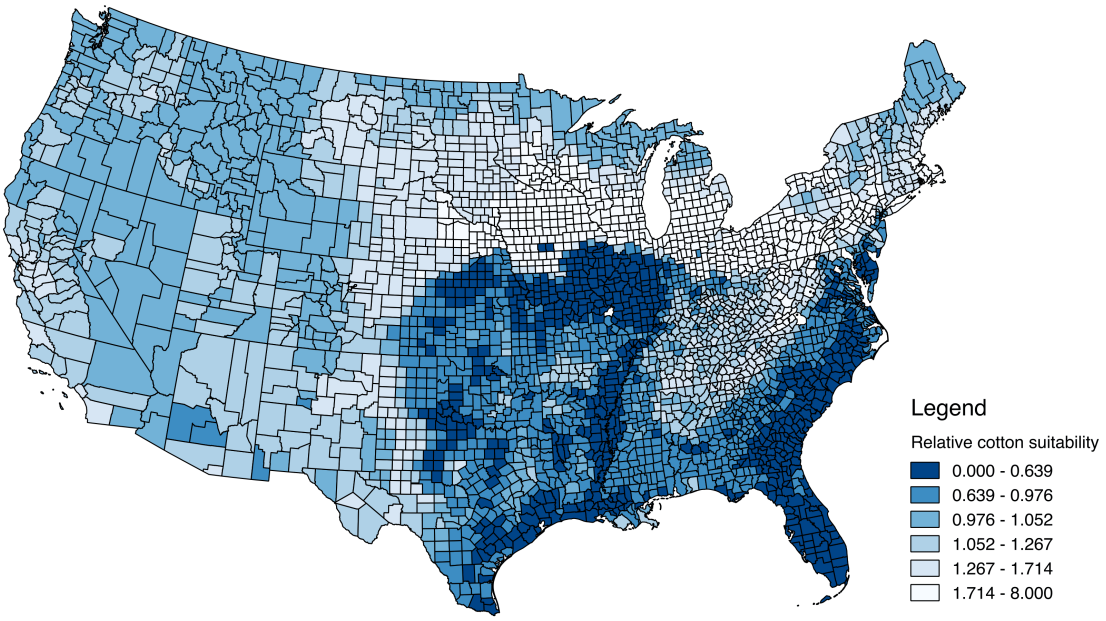
Table B.6: Social Networks

	Labour Force Participation	Log Occupation Income Score
	(1)	(2)
Relative Cotton Production 1840 × female	.069***	.141***
	(.009)	(.017)
Number of Female Slaves 1860 × female	.076***	.090***
	(.019)	(.022)
Female	-.817***	-.616***
	(.023)	(.046)
Individual Controls	Y	Y
County FE	Y	Y
State FE	Y	Y
Industry FE	Y	Y
Counties	747	746
N	186,563	134,798
R-Squared	0.4886	0.2547

Notes: Standard errors clustered at the county level.***p>0.01 **p>0.05 *p>0.10. Columns 2 and 3 include individuals of age between 25 and 54. The regression estimates in column 1 are at the county level and refer to mixed race couples where the wife is African American and the husband is white. The coefficients of the variable *relative cotton production* correspond to the change in the dependent variable due to a 1 standard deviation change in relative cotton production in 1840 or in relative suitability. Data from the 1880 Population Census.

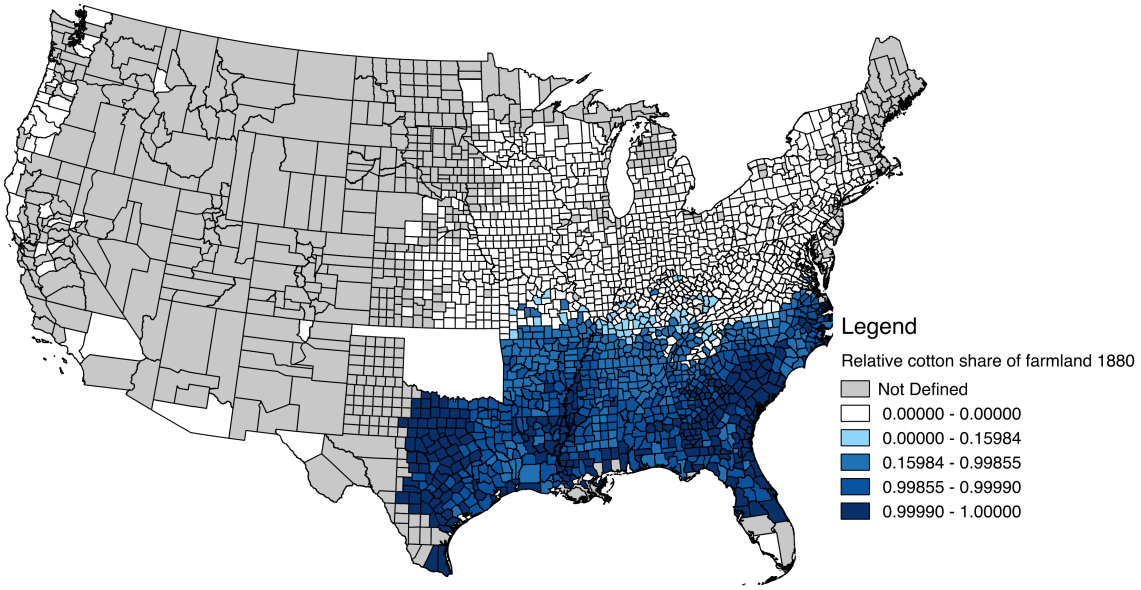
C Figures

Figure C.1: Relative Cotton Suitability



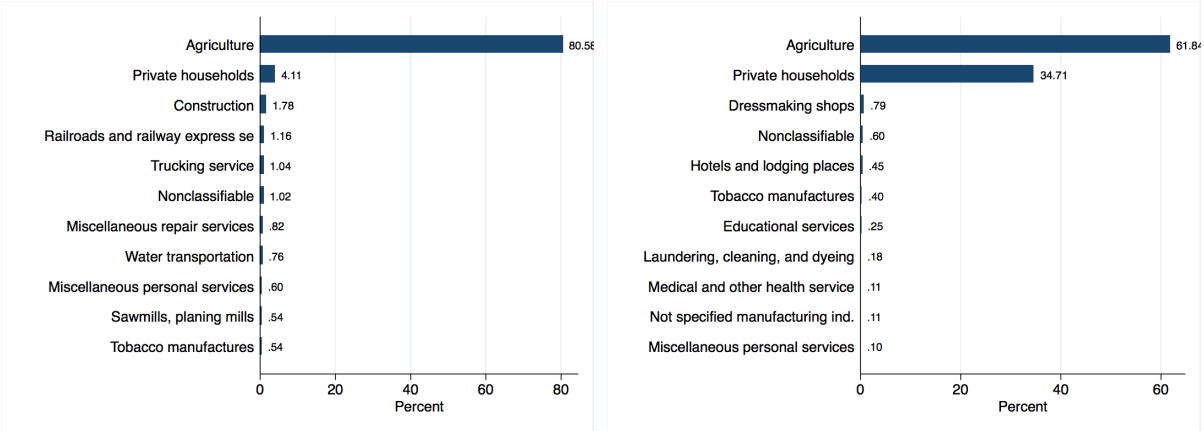
Notes: This map shows the geographic distribution of cotton suitability relative to tobacco by county in the U.S. Source: GAEZ-FAO Crop Suitability Data.

Figure C.2: Relative Cotton Acreage 1880



Notes: This map shows the relative share of cotton to tobacco acreage in 1880, measured in acres. Source: U.S. Agricultural Census 1880.

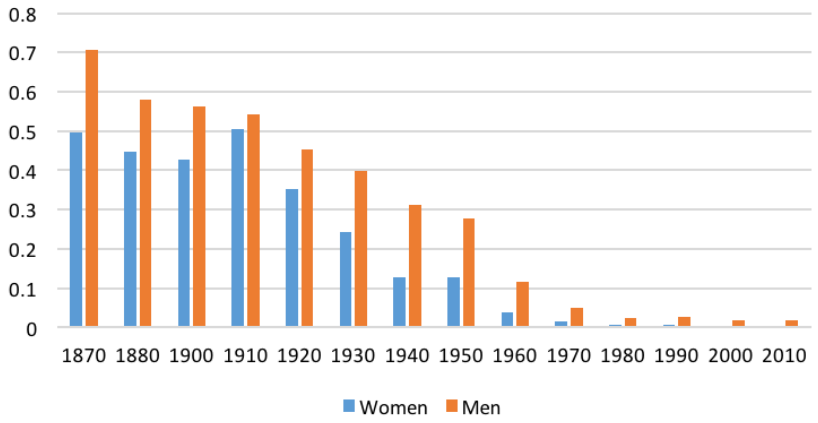
Figure C.3: Main Industries by Gender 1880



(A) Men (B) Women

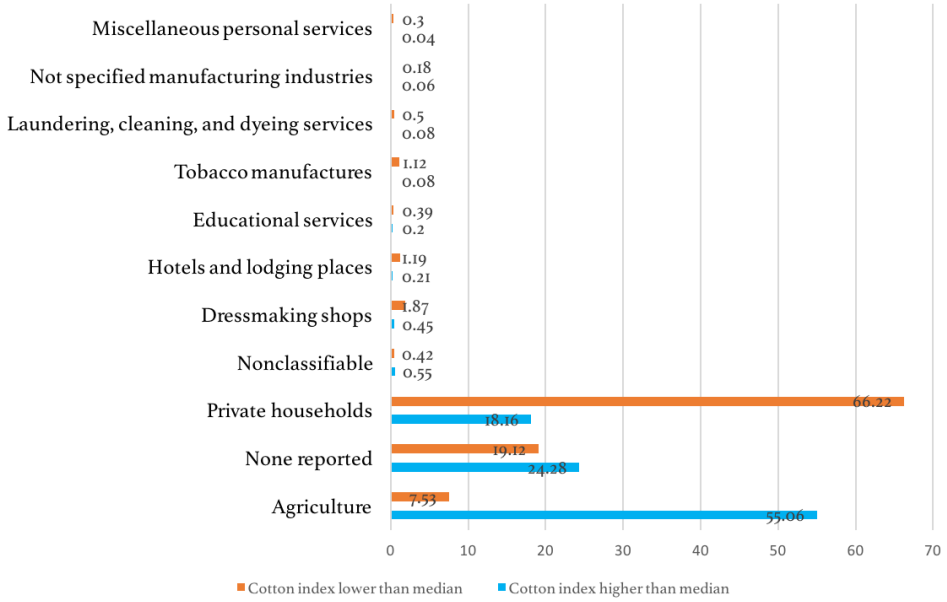
Notes: This graph shows the main 10 industries by employment of African American men (left) and women (right) in 1880. Source: U.S. Population Census 1880.

Figure C.4: Share of African Americans Employed in Agriculture by Gender



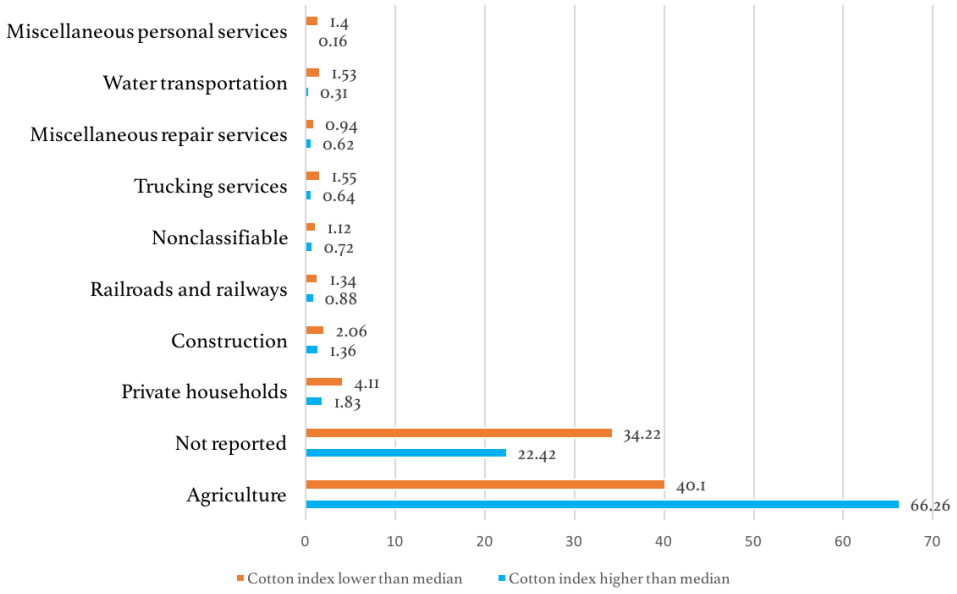
Notes: This graph shows the share of African Americans by gender in the labour force who work in agriculture over time. Source: U.S. Census.

Figure C.5: African American Women's Occupations by Cotton Prevalence, 1880



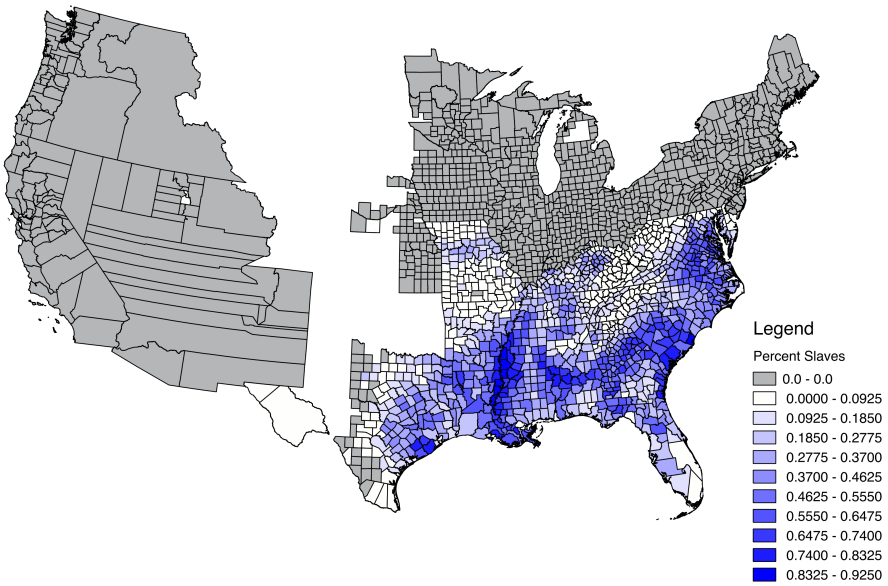
Note: This graph shows the share of African American women in the labour force who work in the top 10 industries in 1880 in counties with cotton index higher than median (blue bars) and below median (orange bars). Source: U.S. Population Census 1880.

Figure C.6: African American Men's Occupations by Cotton Prevalence, 1880

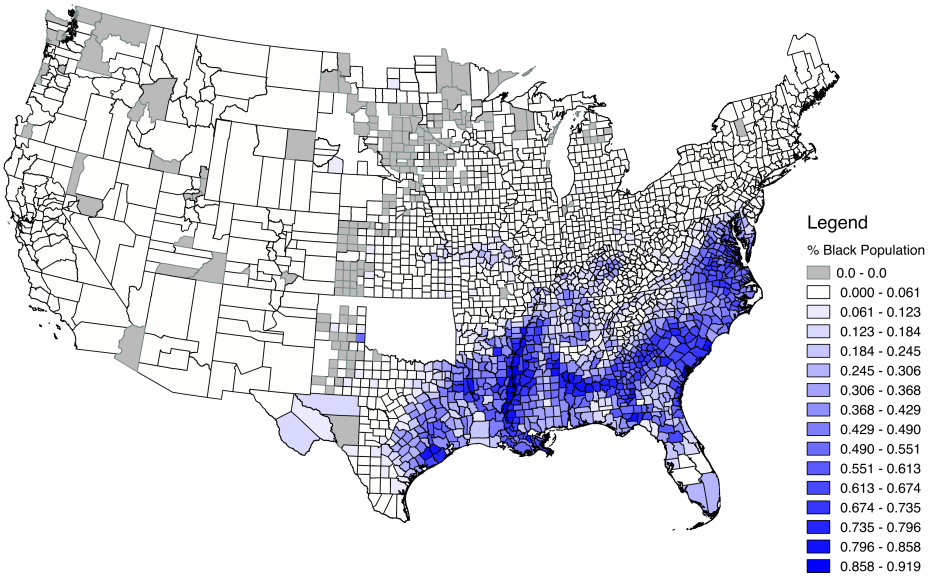


Notes: This graph shows the share of African American men in the labour force who work in the top 10 industries in 1880 in counties with cotton index higher than median (blue bars) and below median (orange bars). Source: U.S. Population Census 1880.

Figure C.7: Slavery 1860 and African Americans 1880 by County



(a) Slaves by County 1860



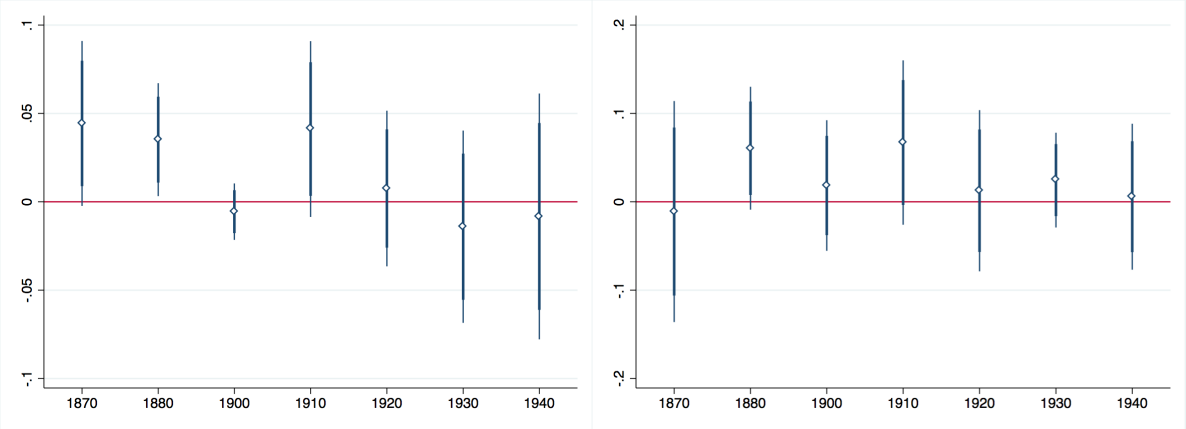
(b) African Americans as Share of Total County Population 1880

Notes: These two maps show the percentage of slaves by county in 1860 and the percentage of African Americans by county in 1880. Source: U.S. Historical Censuses

Figure C.8: Surnames of Cotton Slaveholders and Women in the Labour force by County 1880

Notes: This graph shows the correlation between the number of African American individuals with surnames that match those of cotton slaveholders (divided by its standard deviation) and the ratio of African American females to males in the labour force in 1880. I exclude the top and bottom 1% of both measures. Data source: U.S. census.

Figure C.9: Persistence Including State×Female Fixed Effects

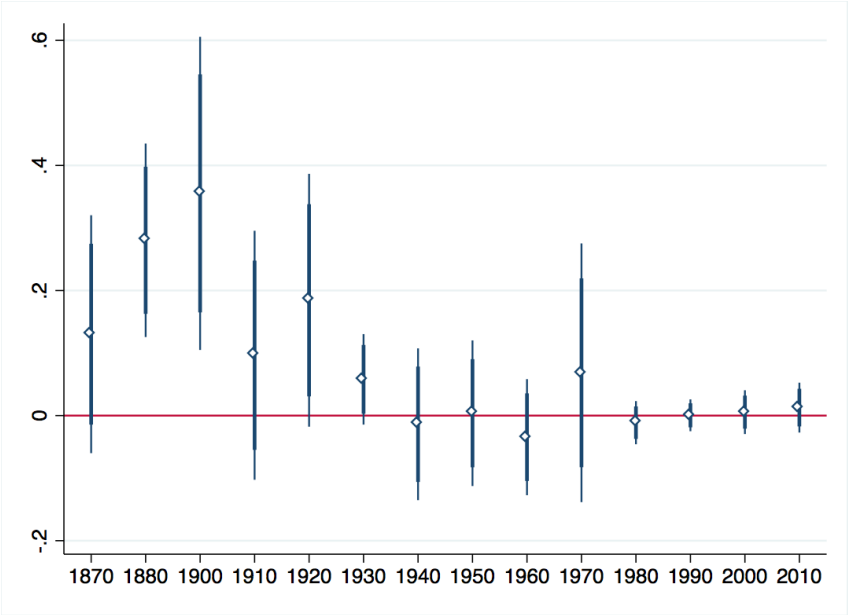


(A) Labour Force Participation (B) Occupation Income Score

Notes: These two graphs show the coefficient plots of regressing labour force participation and occupation income score on the variable $relative\,cotton\,production \times female$ using census data for years 1870-1940, estimated with OLS regressions. The size corresponds to the effect of a one standard deviation change in the dependent variable. The confidence intervals reported are 95 and 99%. I include individual controls, county fixed effects, county characteristics×female and State fixed effects×female. Standard errors are clustered at the county level.

Figure C.10: Persistence - IV Estimates

(a) Labour Force Participation

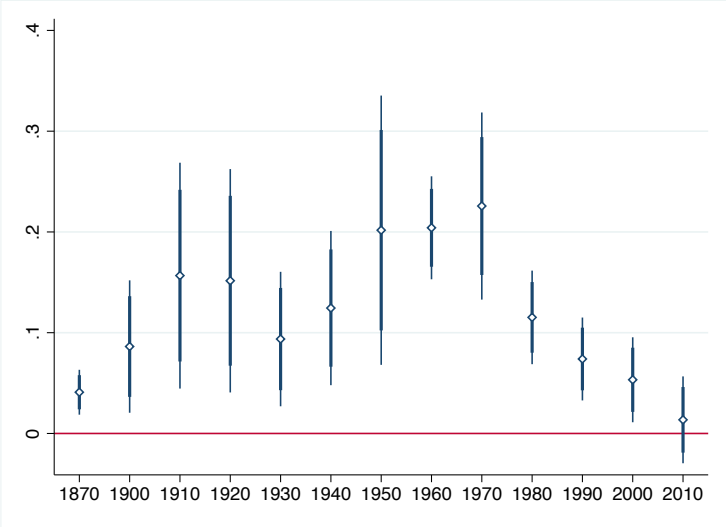


(b) Occupation Income Score

Notes: These two graphs show the coefficient plots of regressing labour force participation and occupation income score on the variable $relative\,cotton\,production \times female$ using census data for years 1870-2010, estimated with instrumental variable regressions. The size corresponds to the effect of a one standard deviation change in the dependent variable. The confidence intervals reported are 95 and 99%. I include individual controls, county fixed effects and county characteristics×female. Standard errors are clustered at the county level.

Figure C.11: Number of Children

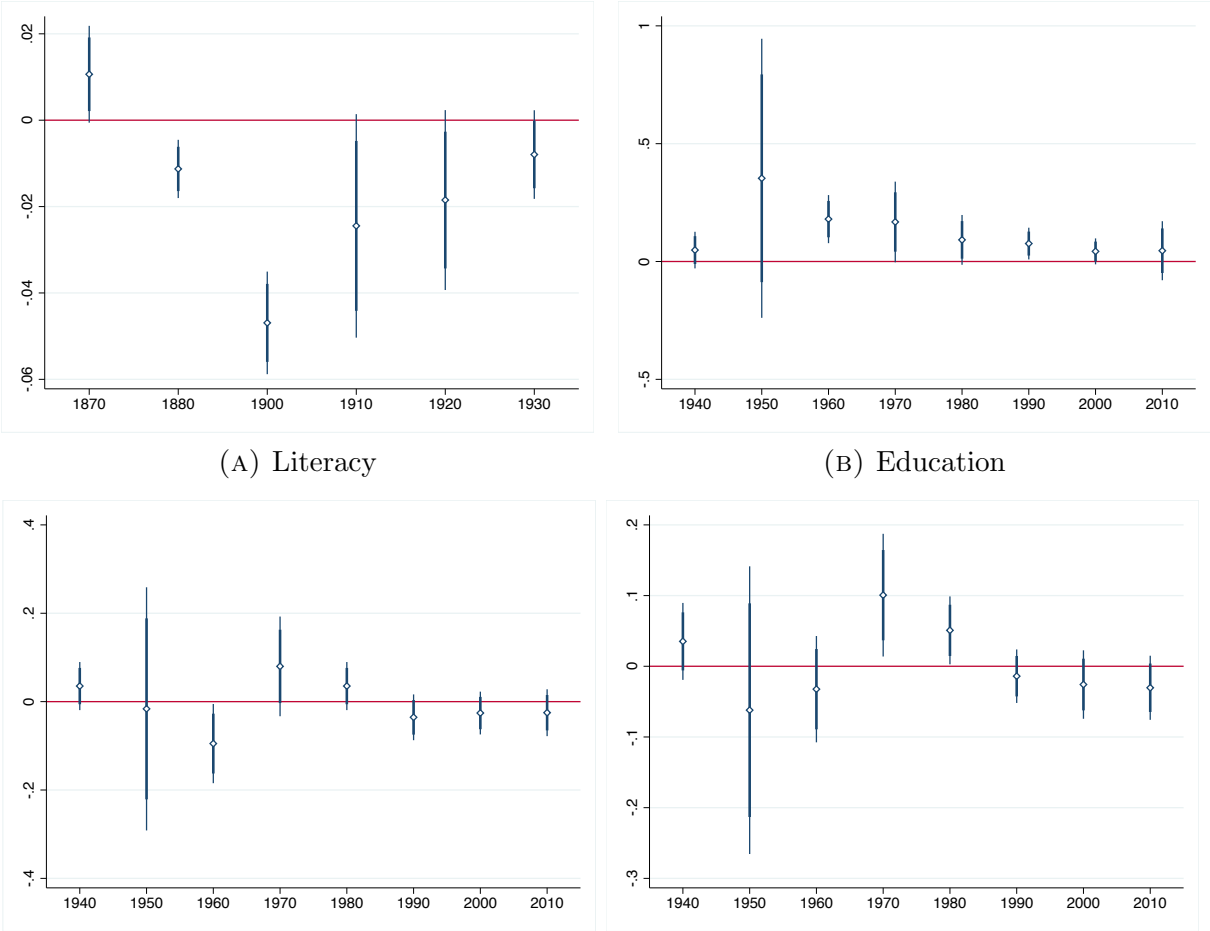
(a) African American Women



(b) African American and White Women

Notes: These two graphs show the coefficient plots of regressing number of children on the variables *relative cotton production* (top) and *relative cotton production*$\times$*black* (bottom) using census data for years 1870-2010. The size corresponds to the effect of a one standard deviation change in the dependent variable. The confidence intervals reported are 95 and 99%. I include individual controls and county fixed effects. Standard errors are clustered at the county level.

Figure C.12: Literacy, Education and Wages



(A) Literacy

(B) Education

(C) Wages

(D) Wages within occupation

Notes: These graphs represent the coefficients of the variable *relative cotton production×female*. The dependent variables are literacy at the top left panel (years 1880-1930), years of education at the top right panel (1940-2010) and log wages (years 1940-2010) in the bottom panels. All regressions include county fixed effects and individual controls, as described in equation (1); panel (d) includes occupation fixed effects. Standard errors are clustered at the county level.